

JPRS 68726

4 March 1977

MACHINE-ASSISTED TRANSLATION IN WEST GERMANY

**DISTRIBUTION STATEMENT A**  
Approved for Public Release  
Distribution Unlimited

20000215 076

U. S. JOINT PUBLICATIONS RESEARCH SERVICE

Reproduced From  
Best Available Copy

REPRODUCED BY  
**NATIONAL TECHNICAL  
INFORMATION SERVICE**  
U. S. DEPARTMENT OF COMMERCE  
SPRINGFIELD, VA. 22161

WEST

EUROPE

JPRS 68726

4 March 1977

MACHINE-ASSISTED TRANSLATION IN WEST GERMANY

<b>BIBLIOGRAPHIC DATA SHEET</b>	1. Report No. JPRS 68726	2.	3. Recipient's Accession No.
	4. Title and Subtitle  MACHINE-ASSISTED TRANSLATION IN WEST GERMANY		5. Report Date 4 March 1977
7. Author(s)		6.	
9. Performing Organization Name and Address Joint Publications Research Service 1000 North Glebe Road Arlington, Virginia 22201		8. Performing Organization Rept. No.	
		10. Project/Task/Work Unit No.	
		11. Contract/Grant No.	
12. Sponsoring Organization Name and Address  As above		13. Type of Report & Period Covered	
		14.	
15. Supplementary Notes			
16. Abstracts  This report contains information on various German systems for machine-assisted translation, automatic composing of dictionaries, automatic querying of terminology data banks and reversible transliteration of cyrillic letters.			
17. Key Words and Document Analysis. 17a. Descriptors  West Germany Machine Translation			
17b. Identifiers/Open-Ended Terms			
17c. COSATI Field/Group 05G			
18. Availability Statement Unlimited Availability Sold by NTIS Springfield, Virginia 22151		19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 110
		20. Security Class (This Page) UNCLASSIFIED	22. Price

4 March 1977

## MACHINE-ASSISTED TRANSLATION IN WEST GERMANY

CONTENTS	PAGE
TEAM: A Program System for the Solution of Terminological and Lexicographical Problems (Karl-Heinz Brinkmann, Eberhard Tanke; SIEMENS DATA PRAXIS, date not given) .....	1
Lexicography With TEAM--Automatic Dictionary Composition (Joachim Schulz; SONDERDRUCK AUS 'DATA REPORT', No 9, 1974) ..	23
Computer Assisted Translation Procedure for Querying a Multilingual Technical Dictionary Stored on Magnetic Tape (J. Schulz; LEBENDE SPRACHEN, No 4, 1975) .....	34
The Use of the Lexicographical Branch of a Data Bank System To Produce a Phraseological Technical Glossary (R. Schmidt, O. Vollnhals; LEBENDE SPRACHEN, Nos 3-4, 1974) ..	43
Reversible Transliteration of Cyrillic Letters (Joachim Schulz; NACHRICHTEN FUER DOKUMENTATION, No 6, 1973) ..	69
Considerations in Setting Up and Operating Terminology Data Banks as a Prerequisite for Machine Assisted Translation (Karl-Heinz Brinkmann; NACHRICHTEN FUER DOKUMENTATION, No 3, 1974) .....	76
Automatic Interrogation of a Terminology Data Bank (Joachim Schulz; NACHRICHTEN FUER DOKUMENTATION, No 1, 1976) ..	89
The Lexicography Information System (Lexis) of the Bundeswehr Language Service (DER SPRACHMITTLER, No 3, 1976) .....	99

TEAM: A PROGRAM SYSTEM FOR THE SOLUTION OF TERMINOLOGICAL AND  
LEXICOGRAPHICAL PROBLEMS

Munich SIEMENS DATA PRAXIS in German [date not given] pp 1-18

[Description of the TEAM program system by Karl-Heinz Brinkmann and Eberhard Tanke: "TEAM - ein Programmsystem fuer die Loesung terminologischer und lexikographischer Aufgaben"]

[Text] Abstract

The problems of international information exchange are to a large extent also problems of language. They can be solved only if the partners in this information exchange have access to the foreign language vocabulary, including technical terminology in not only all possible forms, but also in a form reflecting the state of the art.

TEAM is a flexible program system developed by Siemens, by means of which the technical vocabulary of any number of languages can be recorded and evaluated. The central feature of the system is a computer stored "dictionary" for both direct access with individual queries as well as for batch processing. The vocabulary, which is recorded in the TEAM program system in correct orthography and in small, individually adressable units, can be prepared, sorted and put out in any desired manner. It can be compiled and printed as technical word indexes in the form of books and lists in one or more languages.

TEAM generates data media for the control of composing machines or high-speed filmsetting equipment, which at the same time contain all of the requisite control criteria for automatic composing.

TEAM can be incorporated into a comprehensive dialogue system and can work in conjunction with other documentation systems, as for example, GOLEM.

Reprinting is gladly permitted provided two voucher copies are sent in to our data processing division and the source is credited: "Siemens Publication Series Data Praxis".

## TABLE OF CONTENTS

1. Introduction
2. The TEAM Program System
3. Input Media
4. Input Format
5. Information Processing
6. Output
  - 6.1. High-speed Printout
  - 6.2. Data Media for Automatic Composing
7. Computer Dictionary
8. Outlook

### 1. INTRODUCTION

Considerable activity and effort is being devoted today to solving problems in the international exchange of information, which result from the so gladly cited information avalanche. Through the application of computer procedures for data documentation and retrieval, any stored information is ready for rapid access. However, if this information is composed in a language foreign to the user, in certain cases it can be worthless to him, solely because the technical vocabulary used in this information is either partially or not at all familiar to him.

The magnitude of the difficulties arising at the "language intersection points" can be gauged from the following comparison figures: according to the data of the Grossen Brockhaus, the German literary language encompasses about 300,000 words, English about 600,000, and the vocabulary of a person of average education is about 50,000 words. There are no data available for the total technical vocabulary from science and engineering. As figures from individual areas show, it is many times greater than the vocabulary of the literary language.

To cite an example: in communications engineering and data processing with their applications areas, a stock of far over one million technical expressions has to be taken into account today. Terminological difficulties result from this plethora, which are multiplied even more, primarily in modern specialist areas, through the short lifetime of designations in these

specialties and the at least quantitatively inadequate technical terminology standardization in the language areas.

Apart from a few starts at solving these problems in relatively small fields<sup>1</sup>, systems for the accelerated preparation of state of the art technical vocabulary collections are generally not employed. The methods of composing technical dictionaries hardly take into account the increasing flood of information. Too much time lapses between the appearance of new foreign language expressions until these terms appear in technical dictionaries with the appropriate data.

In order to improve the level of efficiency in international exchange, the readers of foreign language information must be offered a state of the art technical vocabulary in any requisite form by means of new procedures. This claim should be taken up by all internationally organized, private and public undertakings, official offices, administrations and professional associations, as well as individual persons such as scientists, engineers and technical translators. It follows from this that technical vocabulary collections are required in diverse compilations and language combinations in book and list form, as well as in the form of "computer dictionaries" which can be interrogated directly.

The totality of the interrelated problems, among others the automation and thereby the rationalization and the acceleration of the production of technical dictionaries, can be solved in an ideal fashion by means of the Siemens program system, TEAM<sup>2</sup>. As a structural component of a dialogue system, the computer dictionary can be interrogated directly via video display terminals or teletypewriters.

## 2. THE TEAM PROGRAM SYSTEM

In the broadest sense of the word, the technical dictionary in this system is primarily a terminology data bank built up through the data processing system (DVA). In addition to technical expressions, it also contains additional information such as definitions, textual examples and source citations. Depending on the application's purpose, either magnetic cards, magnetic discs or magnetic tapes are employed as storage media. The content of this technical vocabulary store can be called up in any order and in any scope desired, and fed out through peripheral equipment. The possibilities in this case extend from the individual query to issuing single and multi-language alphabetical and systematic technical vocabulary

---

<sup>1</sup> DICAUTOM (*dictionnaire automatique*) of the Terminology Office of the European Association for Coal and Steel, in conjunction with the *Linguistique Automatique Appliquée* of the Free University in Brussels.

Procedure for computer assisted translation of the translation service of the Bundeswehr.

<sup>2</sup> Terminology Acquisition and Evaluation Method.

collections and glossaries. The working aids offered by the data processing system are, however, not utilized completely for just the lexicographical, but also for the terminological side of the work. The requisite work sequences for the preparation of technical dictionaries, such as the compilation, ordering, comparison and systematization of the vocabulary, the correction of word locations, the preparation of discussion manuscripts, etc., are taken over by the computer, just as the selection and re-ordering, the exclusion of duplicates, synonym handling and all other problems which come up in dictionary preparation and publication. Shown in the simplified sequential plan of Figure 1 are the important steps in the preparation of technical dictionaries and the work sequences to be executed by the data processing system. The data flow chart in the appendix supplies additional information.

The TEAM system consists of a series of modularly built-up programs, which break down the input, processing and output of the word material into individual, easily comprehensible sections which are independent of each other.

The programs are written in assembler language and require a Siemens 4004/35 data processing system, or a larger model with a minimum central memory capacity of 65, preferably 131 Kilobytes (KB).



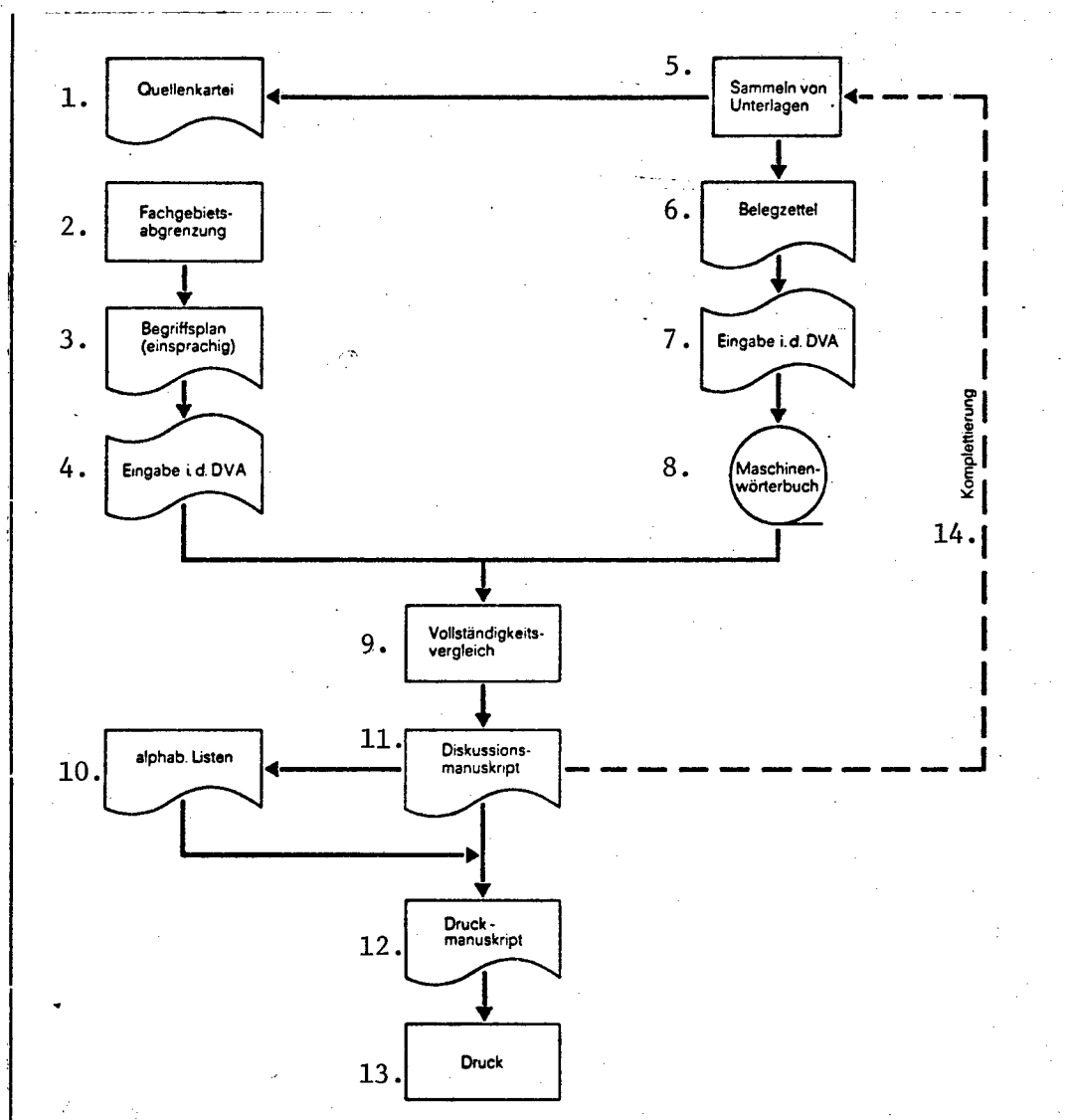


Figure 1. Producing multilingual, systematic dictionaries with data processing systems

Key: 1. Source card index; 2. Technical area limiting; 3. Conceptual plan (monolingual); 4. Data processing system input; 5. Data information collection; 6. Documentation card; 7. Data processing system input; 8. Computer dictionary; 9. Completeness comparison; 10. Alphabetical lists; 11. Discussion manuscript; 12. Printing manuscript; 13. Printout; 14. Completing.

### 3. INPUT MEDIA

Considered as input media are punched cards, perforated tapes or directly recorded magnetic tapes. Because punched cards are too unwieldy and troublesome, as will be yet shown in more detail, and because magnetic tape recorders with the requisite symbol capacity require a relatively high outlay, perforated tapes are preferred for the version of the TEAM program system employed at the present time.

```

$$$$$
03  d
04  0968
05  0183
06  E61
12  DIN 50100V
22  DIN 50100V
32  DIN 50100V
52  DIN 50100V
99</
00  AK2300
10  Dauerfestigkeit im Druck-Schwellbereich
20  fatigue strength under pulsating compressive stress
26  fatigue strength under pulsating>
26  fatigue strength under oscillating compressive stress
>></
00  AK2300
10  Dauerfestigkeit im dr>
10  Dauerfestigkeit im Druck-Schwellbereich
20  fatigue strength under pulsating compressive stress
26  fatigue strength under oscillating compressive stress;
    fatigue strength under pulsating>
    fatigue strength under fluctuating compressive stress
30  limite de fatigue en zone des efforts ondulés par
    compression
36  limite de fatigue en zone des efforts ondulés par
    compression
50  predel ustalosti v oblasti znakopostojannoj cikličeskoj
    nagruzki pri sžatii
99</
k0  AK2300
36  limite de fatigue en zone des efforts répétés par
    compression
99</

```

Figure 2. Input form (106 Teletype) with lead-in (\$\$\$\$\$, lines 03 - 52), brief entries, synonyms (lines 26 and 36) and corrections:

- line erase = >
- word location erase = >><
- subsequent erase and re-input of information (k0 AK2300, line 36, replaces line 36 in the previous entry 00 AK2300).

The 106 Teletype developed by Siemens for documentation purposes is suitable as the input unit. It supplies a six track perforated tape and has a capacity of 116 symbols. A number of requirements which are to be placed on a dictionary program system under all circumstances can be met by it:

--true orthography with upper and lower case letters, umlauts and diacritical marks;

--representation of these symbols by the simplest means, i.e., without troublesome manipulations of the input unit, as is the case with a simple typewriter;

--instant writing of an input form (Figure 2) in true orthography, simultaneously with the production of the input data media (perforated tape). The form protocol should be readable by everyone and require no interpreting of special symbols or special symbol sequences (important for proofreading).

--the capability of correction at any point in time during and after the initial input.

There are thus no special or phantom symbols when using this input unit, as are necessary in the case of punched cards for example, if one wants to distinguish between upper case and lower case letters, as well as umlauts and other letters with diacritical marks. A further advantage of the perforated tape over the punched cards in the input of text is having the ability to continually record data, i.e., not have to repeat any data. On the other hand, this is required when working with punched card sequences, if the length of the information to be recorded amounts to more than 80 written positions.

#### Der große Duden - Band 1 - Rechtschreibung:

Äll/gäu

Äl/les\_bes/ser/wis/ser

Äl/li:anz

#### Gerhard Wahrig - Deutsches Wörterbuch:

'auf.zeich/nen

ünung

Auf.wie.ge'lei

#### Webster's Third New International Dictionary:

ab'er.ra'tion

ab'sent-mind'ed

Figure 3. Representation of stress and syllable separation symbols with the 106 Teletype.

With the teletype used here, letters and diacritical marks can be written separately as in the case of a normal typewriter. Through any combination of upper or lower case letters with about 16 diacritical symbols, one has a supply of nearly 1000 symbols overall at his disposal. Furthermore, different meanings can be assigned to these almost 1000 symbols depending on the nature of the information, and for which fixed written form they are employed. Thus, non-Latin scripts (Greek, Cyrillic, phonetics) can be transliterated without difficulties in accordance with the rules of national and international standards; consequently, they remain quite readable.

Likewise, there are no difficulties in providing the designations with syllable separation and stress marks, and in fact, in any desired form (for example, in accordance with Duden, Wahrig, Webster, etc.).

#### 4. INPUT FORMAT

An input format is used for the input which is graphically clear, and capable of being varied and expanded as desired. It meets the following requirements:

- any number of individually addressable types of information within a word location;
- no limitation on text lengths (Bytes) for any type of information;
- no limitation as regards the number of languages within a word location;
- one-time writing of all information which remains unchanged in a sequence of any number of word locations;
- manifold possibilities for making corrections, for example, for breaking off and erasing lines, breaking off and erasing entries (word locations), subsequent supplementation or erasing of word locations or parts of word locations on the same or any other perforated tape;
- practically no lengthwise limitation on word locations.

The subdivision of the input data into individually addressable "information categories" makes simple access possible to the specific information of interest in a word location. It also makes it possible to assign different meanings to the individual symbols in the different information categories, such as have already been indicated under "Input Media".

The information categories take into account all information which could be of interest in relationship to a word location, thus, besides the word or the word sequence (designation) and the associated synonyms. For example, the nature of the word:

- source (location),
- technical field,
- association with a particular piece of equipment or system,
- references to illustrations,
- input date,

input operator,  
quality data (preferred, permitted or impermissible designations,  
or the like).

Since the number of print positions available for the individual types of information is not limited, definitions and contextual examples can also be recorded for the individual word locations, as well as for the individual languages within word locations.

Should additional information be desired, for example, that of linguistic interest, the input format can be expanded without difficulty. If no definitions or contextual examples are recorded onto punched tapes, references to the location of such information, or to a microfilm index which contains this information, can be provided in the input format.

## 5. INFORMATION PROCESSING

A number of programs are available for processing information recorded on perforated tapes (see the appendix).

The perforated tapes are transferred to magnetic tapes by means of the LOMA program component. The input data are checked for formal correctness, normalized, and if necessary, erased with the issuance of an error voucher. Brief entries are extended to full length standard entries by means of constant information which is recorded only once and supplied in the form of a "lead-in". The error voucher can be put out selectively either via a high-speed printer or on magnetic tape.

The KORR program carries out a multiplicity of functions. Among other things, it executes the corrections of word locations. Above and beyond this, it makes other checks, supplies a report on amendments which have been undertaken as required, and works out some statistical data. A variant of this program (UDNR) brings the stored word stock up to date.

The SESUS program solves some particularly difficult problems. It uses the synonyms taken in by individual word locations to generate its own word locations for these synonyms. This means a considerable facilitating of the input work, during which otherwise the synonyms in the output language would have to be combined (permutation) with all synonyms in the target language (or target languages).

Additionally, in case it is desired, this program breaks compounds down into their members and generates in accordance with the transposition of these members, so-called "transposition entries". By way of example, the program generates the sequence "noun - comma - adjective" from the word sequence "adjective - noun", for example, "automatic exchange" becomes "exchange, automatic".

Furthermore, depending on interest and availability, this program selects certain partial information from the totality of the stored word location information.

By way of example, a selection can be made in accordance with the following considerations:

- Quality of the word location information (preferred, permissible or impermissible designations);
- Languages (in any combination);
- Technical fields (individually or in combination);
- Sources;
- Word types;
- Input date (check to see if the vocabulary is up to date!);
- The existence of certain information in any categories, for example, definitions, contextual examples, illustrations or the like.
- Association with a system or piece of equipment.

Besides this "selection", this program also takes over the "sorting", i.e. it sets up a so-called "sorting concept" for the languages, in accordance with which the vocabulary is to be alphabetically sorted. By means of this sorting concept, the vocabulary can be arranged in any sequence and over any number of locations, depending on the desire of the user. Any arbitrary type of sequence, especially as regards the classification of the umlaut letters, and any deviant sorting sequence in other Latin and non-Latin alphabets, can be taken into account without difficulty (special letters in Latin alphabets, as for example, "ß" in German, "ñ" and "ll" in Spanish, etc.; alphabetizing in Greek, Cyrillic, etc.).

In forming the sorting concept, inserts in the text which have to remain out of consideration during the sorting (Roman numerals, which the data processing system would interpret as letters; Latin letters and letter sequences within a Cyrillic text, etc.) are either suppressed or changed into a suitable form.

With the enormous quantities of data which are to be recorded, one cannot permit oneself the luxury of checking in each individual case whether a word has already been recorded from another source. Thus the input is undertaken without such a check, and as a consequence there are word locations which are present in many places. One speaks in a general sense of "doublets". A special program (DUBL) is employed to clean them up.

Its main function consists in bringing word entries in which the technical expressions are identical in the individual languages together into one individual word location. In this case, all source and technical field data are retained though, as well as other data of interest, i.e. they are transferred from the word locations to be erased to a permanent word location.

The synonyms located at the original word location are likewise brought together in such a fashion that each synonym is only entered once in the permanent word location.

Example (Data configuration on the magnetic tape):

First word location:	06 E1600; E6500	(technical fields)
	10 Halbleiter diode	(German)
	11 f. [noun]	(part of speech)
	12 WWB WH	(source)
	20 semiconductor diode	(English)
	22 WWB WH	(source)
	26 crystal diode	(English synonym)
Second word location	06 E6500	
	10 Halbleiterdiode	
	11 f.	
	12 DIN 41855E	
	20 semiconductor diode	
	22 DIN 41855E	

The DOUBL [sic] program brings these two word locations together into one individual word location:

```

06 E1600; E6500
10 Halbleiterdiode
11 f.
12 DIN 41855E; WWB WH
20 semiconductor diode
22 DIN 41855E; WWB WH
26 crystal diode

```

Because the doublet comparison also incorporates the already mentioned sorting concepts, in many cases the program is capable of recognizing doublets itself, if certain differences exist in the manner of writing. In such cases as this, and in similarly occurring ones, the program reports a "doublet suspected".

## 6. OUTPUT

As regards the output, the TEAM program system is also extraordinarily flexible. It can work with the most diverse output media and output formats. Specifically suitable programs and program variants are available for this purpose.

### 6.1. High Speed Printer Output

The print programs (DRU1, DRU2) are suitable for output both through conventional high-speed printers, which have only upper case letters or lower case letters at their disposal, as well as through special high speed printers,

the symbol capacity of which takes in upper and lower case letters, as well as diacritical marks. One such "library high speed printer" puts out the

KS0044	entwickeln v. UIT DRAFT E4300	develop v. UIT DRAFT	röveler v. UIT DRAFT 1268 0159	revelar v. UIT DRA	projavljat' v. UIT DRAFT
KS0045	Blende f. UIT DRAFT E4300	diaphragm n. UIT DRAFT	diaphragme (en photostéographie) m. UIT DRAFT 1268 0159	diafragma f. UIT DRAFT	diafragma f. UIT DRAFT
KS0046	Sendeverzerrung f. UIT DRAFT E4300	transmitter distortion UIT DRAFT	distorsion à l'émission UIT DRAFT 1268 0159	deformación en la emisión UIT DRAFT	iskazenie pri peredatse UIT DRAFT
KS0047	einseitige Verzerrung UIT DRAFT unsymmetrische Verzerrung E4300	bias distortion UIT DRAFT dissymmetric distortion	distorsion blaise UIT DRAFT distorsion dysymétrique 1268 0159	deformación asimétrica UIT DRAFT deformación disimétrica	preobladanie n. UIT DRAFT
	DEUTSCH	ENGLISCH	FRANZ.	SPAN.	RUSS.
KS0048	Verteiler m. UIT DRAFT E4300	distributor n. UIT DRAFT	distributeur m. UIT DRAFT 1268 0159	distribuidor m. UIT DRAFT	raspredelitel' m. UIT DRAFT
KS0049	Löschung (Lochner) f. UIT DRAFT E4300	erasure of errors (perforator) UIT DRAFT	effacement des erreurs (perforateurs) UIT DRAFT 1268 0159	rectificación de errores (perforadores) UIT DRAFT	zaboj ořibok UIT DRAFT

Figure 4. Discussion manuscript, five languages, put out on the library high speed printer (numerically sorted, with page alternation).

stored vocabulary in the form required for the linguistic data processing, and especially for the lexicographical work. The programs also permit an output to magnetic tape, in order to make possible the economically advantageous use of off-line high-speed printers.

Printout via conventional high speed printers, which permit only either upper case or lower case writing, and whose store of special symbols is limited and includes no diacritical marks, is suited for test purposes and for short lived word lists, with which only well-versed linguists should work. For example, they are not suitable as discussion manuscripts, which are to be read linguistically less well trained technical specialists, or as word indexes, which are to be made available to users with only passive language knowledge.

The particular languages from the high speed printer program are normally arranged one below the other within the word locations. However, there is a variant (DRUN) in which up to five languages can be written horizontally next to each other (Figures 4 and 5).



	DEUTSCH	ENGLISCH	FRANZ.
D10050	BETATEILCHEN N. DIN 25401 EIN ELEKTRON POSITIVER ODER NEGATIVER LADUNG, DAS VON EINEM ATOMKERN ODER ELEMENTARTEILCHEN BEIM RADIOAKTIVEN ZERFALL AUSGESANDT WIRD. E2000	BETA PARTICLE DIN 25401	PARTICULE BETA DIN 25401       0369 0183
D10051	BREMSVERMOEGEN, NEUTRONEN- DIN 25401 PRODUKT AUS DEM MITTLEREN LOGARITHMISCHEN ENERGIEDEKREMENT UND DEM MAKROSKOPISCHEN STREUQUERSCHNITT FUEER NEUTRONEN. E2000	SLOWING-DOWN POWER DIN 25401	POUVOIR DE RALENTISSEMENT DIN 25401       0369 0183
D10052	BRENNELEMENT N. DIN 25401 KLEINSTER, SELBSTAENDIGER BAUTEIL, DER KERNBRENNSTOFF ZUR VERWENDUNG IN EINEM REAKTOR ENTHAEHLT. E2000	FUEL ELEMENT DIN 25401	ELEMENT DE COMBUSTIBLE DIN 25401       0369 0183
D10053	BRENNSTOFFHUELLE F. DIN 25401 UNMITTELBAR AUF DEN KERNBRENNSTOFF AUFGEBRACHTE, DICHTER UMHUELLUNG, DIE DIESEN GEGEN EINE CHEMISCH AKTIVE UMGEBUNG SCHUETZT UND DEN AUSTRITT DER WAEHREND DES EINSATZES DES KERNBRENNSTOFFS GEBILDETEN SPALTPRODUKTE VERHINDERT. E2000	CLADDING N. DIN 25401 CLAD;N.	GAINE F. DIN 25401                0369 0183
D10054	BRENNSTOFFHUELSE F. DIN 25401 DICHTER UND FORMFESTE HUELSE ZUR AUFNAHME DES	CAN N. DIN 25401	GAINE F. DIN 25401

Figure 5. Discussion manuscript, three languages, with definitions, put out via a conventional high printer (arranged alphabetically according to the German designations).

Applicable to all of the output programs described up to this point, as well as for those which are yet to be described, is the fact that any information can be suppressed in the individual word locations on request.

Furthermore, there is the option of being able to produce alphabetical keyword indexes for multilingual alphabetical and systematic dictionaries using suitable variants of this program.

## 6.2. Data Media for Automatic Composing

The DIGIA and SATZ programs handle the preparation of the vocabulary for composition and its transfer to suitable data media (magnetic tape or perforated tape). In this case, all relevant national and international standards and draft standards can be taken into account. The breakdown of the data recorded and processed with the TEAM program system into the smallest, individually addressable information units, makes it possible for these programs to provide all the criteria requisite for automatic composition, so that, among others, the following requirements are met:

- Any printing style format;
- Automatic selection of type fonts, depending on the information to be presented;
- Automatic line and page feed;
- Automatic column and page numeration;
- Automatic maintenance of the column and page lengths provided, as well as of the line or column and page widths;
- Composition without justification, or when the appropriate syllable separation program is engaged, syllable separation and automatic margin equalization;
- Automatic generation of current column headings;
- Automatic back-transliteration of non-Latin scripts.

The DIGIS program produces special data media for the control of the DIGISET photo-composing system of the Dr.-Ing. Rudolf Hell Company, Kiel (Figures 6 to 11).

DIN 25401  
Grobsteuerelement n Steuerelement  
zur Grobeinstellung der  
Reaktivität eines Kernreaktors  
oder zum Ändern der  
Flußdichte Verteilung.  
DIN 25401E  
Größe, kritische  
Mindestabmessungen einer  
Brennstoffanordnung, die bei  
bestimmter geometrischer  
Anordnung und  
Materialzusammensetzung  
kritisch gemacht werden kann.  
DIN 25401

Figure 6. Monolingual definition dictionary (DIGISET).

**Gruppenübergangsquerschnitt  $m$**  Für die Energiegruppenstruktur charakteristischer mittlerer Wirkungsquerschnitt, der den Übergang von einer zu einer anderen Gruppe beschreibt.  
DIN 25401E

**Gruppenverlustquerschnitt  $m$**  Für die Energiegruppe charakteristischer mittlerer Wirkungsquerschnitt, der den Verlust von Neutronen aus dieser Gruppe durch alle Vorgänge beschreibt.  
DIN 25401E

**Halbwertszeit  $t$**  Zeit, in der im Mittel die Hälfte der ursprünglich vorhandenen Atome eines Radionuklids sich umgewandelt, bzw. bei Isomeren in den Grundzustand übergeht.  
DIN 25401E

**Halbwertszeit, biologische** Die Zeit, in der die Hälfte einer bestimmten Substanz aus einem biologischen System durch biologische Vorgänge ausgeschieden wird, wobei angenommen wird, daß die Ausscheidung exponentiell mit der Zeit verläuft.  
DIN 25401

**Halbwertszeit, effektive** Die Zeit, in der die Menge eines

Figure 6 [Continued]

Monolingual definition dictionary  
(DIGISET)

## 7. COMPUTER DICTIONARY

As already noted, the vocabulary taken in by the TEAM program system can also be made available for direct query via a data viewing terminal (display) or page printer and teletype. In this form, it can also offer the users a comprehensive information system, with which they can conduct a dialogue. An information system user, who in studying any information comes up against a foreign language expression he does not know, can have the translation of this expression also put out by the system. On the other hand, a technical language worker or translator, who is engaged in working with the stored vocabulary, can have pertinent technical information fed out through the information system.

## 8. OUTLOOK

The TEAM program system was designed and placed in service by the foreign language service of the data processing and communications engineering divisions of the Siemens Company. As an open system, it makes room for all

<b>gleichgewichtiger Code</b> [E4201, E441, E449] fixed-count code, fixed-ratio code	dollar;
<b>Gleichheitsglied</b> <i>n</i> [E423] equality circuit, equality unit	- <b>Druckaufbereitungszeichen</b> (COBOL) [E425] floating report sign (COBOL);
<b>Gleichheitszeichen</b> <i>n</i> [E11, E4201] equal sign	- <b>Währungszeichen</b> (COBOL) [E425] float dollar sign (COBOL), floating dollar
<b>Gleichlaufprüfung</b> <i>f</i> [E4281] synchronism check, synchronous check	<b>Gleitkomma</b> <i>n</i> [E4201, E425] floating point
<b>Gleichlaufprüfung-Kontrollwort</b> <i>n</i> [E4281] sync check word, synchronous check word	<b>Gleitkomma-Addition</b> <i>f</i> [E424, E425] floating add, floating point addition
<b>Gleichlaufschwankung</b> <i>f</i> [E4114] flutter <i>n</i>	<b>Gleitkomma-Addition absolut</b> [E425] floating add absolute
<b>gleichmäßige Konvergenz</b> [E11, E4201] uniform convergence	<b>Gleitkommaarithmetik</b> <i>f</i> [E4201, E425] floating decimal arithmetic, floating point arithmetic
<b>Gleichung</b> <i>f</i> [E11, E4201] equation <i>n</i> ; - <b>dritten Grades</b> [E11, E4201] cubic equation; - <b>höherer Ordnung</b> [E11, E4201] higher order equation; <b>algebraische</b> - [E11, E4201] algebraic equation; <b>biquadratische</b> - [E11, E4201] biquadratic equation; <b>homogene</b> - [E11, E4201] homogeneous equation; <b>simultane lineare</b> - [E11, E4201] simultaneous linear equation	<b>Gleitkommabefehl</b> <i>m</i> [E425] floating-point instruction
<b>Gleichungsaauflöser</b> <i>m</i> [E426] equation solver	<b>Gleitkommadarstellung</b> <i>f</i> [E4201] floating-point representation, variable-point representation
<b>gleichzeitig</b> <i>adj</i> [E422] simultaneous <i>adj</i> , concurrent <i>adj</i>	<b>Gleitkomma-Division</b> <i>f</i> [E424, E425] floating-point division
<b>gleichzeitige Übertragung</b> [E422, E449] simultaneous transmission	<b>Gleitkommaeinrichtung, festverdrahtete</b> [E422] floating-point feature, automatic floating-point feature
<b>gleitende Division</b> [E424, E425] floating divide;	<b>Gleitkommagültigkeit</b> <i>f</i> (Programmaske) [E425] significance mask
- <b>Hauptspeicheradressierung</b> [E425] floating storage addressing	<b>Gleitkommakonstante</b> <i>f</i> [E425] floating-point constant;
- <b>Multiplikation</b> [E424, E425] floating multiply; - <b>Subtraktion</b> [E424, E425] floating subtract;	- <b>doppelter Genauigkeit</b> [E425] double-precision floating point constant; - <b>einfacher Genauigkeit</b> [E425] single-precision floating point constant
- <b>Zeichenfolge</b> [E425] floating string	<b>Gleitkomma-Konstante mit erweiterter Mantissenlänge</b> [E425] long-precision floating point constant
<b>gleitender Divisionsrest</b> [E424, E425] floating divide remainder	<b>Gleitkommakonstante mit kurzer Mantissenlänge</b> [E425] short precision floating point constant
<b>gleitendes Dollarzeichen</b> [E425] float dollar sign (COBOL), floating	<b>Gleitkomma-Mantissenlänge</b> <i>f</i> [E424, E425] floating-point mantissa length

Figure 7. Dual Language Dictionary (DIGISET)<sup>1</sup>

<sup>1</sup> "Woerterbuch der Datenverarbeitung Deutsch-Englisch/Data Processing Dictionary, English/German", Siemens Aktiengesellschaft, Berlin, Munich, 1970.

**disintegration**

**disintegration energy**

*F*: énergie de désintégration

*D*: Zerfallsenergie *f*

*DI 1017*

**disintegration rate**

*F*: taux de désintégration

*D*: Zerfallsrate *f*

*DI 1018*

**dispersion fuel**

*F*: combustible en dispersion

*D*: Brennstoff, dispergierter

*DI 0223*

**dose equivalent (radiation protection)**

*F*: équivalent de dose

(radioprotection)

*D*: Äquivalentdosis *f*

(Strahlenschutz)

*DI 0214*

**dual-cycle reactor**

*F*: réacteur à double cycle

*D*: Zweikreisreaktor *m*

*DI 1131*

**effective half-life**

*F*: demi-vie résultante

*D*: Halbwertszeit, effektive

*DI 0014*

**electric-power reactor**

*F*: réacteur de production

d'électricité

*D*: Stromerzeugungsreaktor *m*

*DI 1116*

**emergency dose**

*F*: dose d'urgence

*D*: Notstandsäquivalentdosis *f*

*DI 1001*

**enriched fuel**

*F*: combustible enrichi

*D*: Brennstoff, angereicherter

*DI 0222*

**enriched material**

*F*: matière enrichie

*D*: Material, angereichertes

*DI 0087*

**enriched-uranium reactor**

*F*: réacteur à uranium enrichi

*D*: Reaktor, angereicherter

*DI 1019*

**enrichment *n* (process)**

Figure 8.

Trilingual Dictionary (DIGISET)

Deichrolle 10/305	Delanium 16/578
Deichschäden 10/305	Delbourg-Zahl 3/149
Deichscharte, Deichschaart 10/340	Delegierung 15/105
Deichschau 10/305	Deli-Kupplung 1/266, 12/289
Deichschleuse, Deichsiel 11/402	Delmag-Frosch 10/307
Deichschutz 10/305	Delon-Greinacher-Schaltung 6/503
Deichschutzwerke 10/306	Delon-Schaltung 2/225, 2/348
Deichselhubwagen 15/187	Delta 10/307
Deichselkraftschreiber 12/132	Deltaausbau der Donau 10/308
Deichselsteuerung 10/306	Delta-HD-Bronze 3/149
Deichsicherungswerke 10/306	Delta-Metall 3/149
Deichstatut 10/304	Delta-Motor 1/682
Deichtreppe 10/306	Delta-Operator 2/542
Deichübernahme 10/306	Deltaplan, holländischer 10/308
Deichunterhaltung 10/303	Delta-Rakete 12/673
Deichverband 10/306	Delta-Ringdichtung 16/203
Deichverlauf 10/307	Delta-Stütz-Isolator 6/573
Deichverlegung 10/307	Demag-Onia-Gegi-Verfahren 4/390
Deichversinkung 10/307	Demi-Alpakka 3/149
Deichverteidigung 10/307	Demodulator 2/73, 13/158
Deichverwaltung 10/306	Demonstrationsokular 13/158
Deichvorland 10/102	den 3/149
Deichzubehör 10/307	Dendrit 3/429
Deichzug 10/307	Dendriten 16/88
Deichzuweg 10/307	Dengeln 8/311
Deighton-Rohr 6/197	Denier 3/149
Deionat 6/197	Densitometer 13/158
dek 3/149	Dental-Anästhesie-Apparate 13/158
Deka 1/69, 2/73	Dental-Bohrer 13/158
Dekadenkondensator 2/406	Dental-Bohrhandstück 13/158
Dekadenwahl 13/427, 14/22	Dental-Bohrmaschine 13/159
Dekadenwiderstand 2/73	Dental-Einheit 13/159
Dekadenzählröhre 13/405	Dentaleinrichtungen 13/159
Dekamired 14/120	Dental-Legierungen 3/149
Dekanter 16/593	Dental-Operationsstuhl 13/159
Dekapieren 3/149, 5/122, 8/324	Dental-Schleifer 13/159
Dekatieren 8/113	Dental-Technikbohrmaschine 13/159
Dekatiermaschine 8/113	Dental-Turbine 13/160
Dekaturechtheit 3/149	Denver-Fahrenwald-Klassierer 4/42
Deklination 2/73, 4/177, 6/197, 10/71, 13/438	Denver-Gold-Jig 4/501
Deklination, magnetische 11/86	Departmental organisation 15/105
Deklinationskreisel 13/158	
Deklinatorium 4/143, 13/158	
Dekomposition eines Netzplanes 15/296	

Figure 9. Two columns out of the register for "Lueger - Lexikon der Technik" ["Lueger - Engineering Lexicon].

Dekompositionsmethode der linearen Planung 15/285	Departure Curves 4/646
Dekompression 6/40	Dephlegmation 16/495
Dekontamination 10/40, 10/378	Dephlegmator 1/69, 4/143, 16/89
Dekontaminationsfaktor 10/378, 10/448	Displacement 12/716
Dekontaminierung 6/348	Displacement-Ruder 12/502
Dekontaminierung von Abluft 16/13	Depolarisation, optische 13/160
Dekontaminierung von Abwässern 16/88	Depolarisationselektrode 2/74
Dekrement, logarithmisches 1/505, 2/73	Depolarisator 2/74, 2/406
Dekupiersäge 8/113	Depolymerisation 3/151
Delaborne-Prisma 14/294	Deponie von Industrierückständen 16/89
Delalots Legierung 3/149	Deposit attack 3/151
Delaminieren 13/158	Depression 4/143
	Depressionsbestimmung in Schächten 4/593
	Depressionsmesser 4/143
	Depressionstelemeter 13/440

Figure 9. [Continued]: Two columns out of the register for "Lueger - Lexikon der Technik" ["Lueger - Engineering Lexicon"].

conceivable developments and applications. For example, a provision is made for its incorporation into a comprehensive dialog system in which the user works directly with the data processing system and the memory can be continually updated from the results of the work. For procedures in computer translation, the vocabulary can be represented in any desired form within the framework of the format being used. All possible additional designations and supplements can be stored and retrieved in any combinations. Thus, for example, roots of words can be given along with the written out designations, as well as grammatical and semantic functions. By tying into the GOLEM information system, there is now the capability of remote interrogation (TEGO interface program). This branch will be further developed, primarily with the goal of making available to large circles of users quite large data banks with mixed information and technical vocabulary, and long additional textual information as well as illustrations. Created thereby are the prerequisites for the intensification of the information exchange between companies, associations and administrative authorities, and therewith also for the optimization of terminological and dictionary work.

## F

- Feinsteuerelement** *n*  
*E:* control member, fine  
*R:* элемент точного управления  
*DI0139*
- Fluß** *m*  
*E:* flux *n*  
*R:* поток *m* *DI0233*
- 2200 m/s-Fluß**  
*E:* 2200 meter per second flux density  
*R:* плотность потока *v* 2200 м/сек.  
*DI0234*
- Flußdichte, Teilchen-**  
*E:* particle flux density  
*R:* плотность потока частиц *DI0235*
- Forschungsreaktor** *m*  
*E:* research reactor  
*R:* исследовательский реактор *DI1037*

## G

- Gammastrahlung** *f*  
*E:* gamma radiation  
*R:* гамма-излучение *DI0069*
- Gammastrahlung, prompte**  
*E:* prompt gamma radiation  
*R:* мгновенное гамма-излучение *DI0012*
- Generationsdauer** *f*  
*E:* generation time  
*R:* время генерации *DI0140*
- Gleichung, kritische**  
*E:* critical equation  
*R:* критическое уравнение *DI0070*
- grau** *adj* (Reaktortechnik)  
*E:* gray *adj* (reactor technology)  
*R:* серый *adj* (технология реакторов)  
*DI0236*
- Grenze, extrapolierte**  
*E:* extrapolated boundary  
*R:* экстраполированная граница  
*DI0071*
- Grobsteuerelement** *n*  
*E:* control member, coarse  
*R:* элемент грубого управления  
*DI0141*

Figure 10. Trilingual Dictionary  
 (With initialization - DIGISET)<sup>1</sup>

- считывания подлинников,  
 считывающее устройство  
 документов
- character reader** [E4101]  
*D:* Zeichenleser *m*  
*R:* считывающее устройство знаков
- character recognition** [E4281]  
*D:* automatische Zeichenerkennung  
*R:* распознавание знаков, опознавание знаков
- character representation** (COBOL) [E425]  
*D:* Zeichendarstellung *f* (COBOL)  
*R:* изображение знаков, представление знаков
- character set** [E4113, E42]  
*D:* Zeichenvorrat *m*, Zeichenmenge *f*  
*R:* запас знаков, комплект знаков, количество знаков
- character spacing** [E4113, E4281]  
*D:* Zeichenabstand *m*, Typenabstand *m*, Zeichenmittenabstand *m*  
*R:* интервал между знаками
- character, check** [E412, E441, E449]  
*D:* Prüfzeichen *n*, Kontrollzeichen *n*  
*R:* контрольный знак
- character, erase** [E4111, E449]  
*D:* Irrungszeichen *n*, Löszeichen *n*  
*R:* знак ошибки
- character, escape** (ESC) [E425]  
*D:* Umschaltzeichen *n*, Steuerzeichen zur Codeerweiterung  
*R:* сигнал переключения, знак переключения
- character, fill** [E42]  
*D:* Füllzeichen *n*  
*R:* заполняющий знак
- character, format effector** (FE) [E425]  
*D:* Formatsteuerzeichen *n*  
*R:* знак управления форматом

- character, layout** [E425]  
*D:* Formatsteuerzeichen *n*  
*R:* знак управления форматом
- character, lower case** [E411, E4113]  
*D:* Kleinbuchstabe *m*  
*R:* строчная буква
- character, upper case** [E4113, E42, E4281]  
*D:* Großbuchstabe *m*  
*R:* заглавная буква
- character-at-a-time printer** [E4113]  
*D:* Zeichendrucker *m*, Buchstabendrucker *m*  
*R:* печатающее устройство знаков, буквопечатающее устройство
- characteristic** *n* [E4201]  
*D:* Charakteristik *f* (Gleitkomma-Exponent)  
*R:* характеристика *f*, показатель *m*
- characteristic curve** [E42]  
*D:* Kennlinie *f*  
*R:* характеристика *f*, характеристическая кривая
- charactron** *n* [E4101, E64]  
*D:* Charaktron *n*  
*R:* характрон *m*
- chart** *n* [E42]  
*D:* Diagramm *n*  
*R:* диаграмма *f*, схема *f*
- chart, dynamic flow** [E425]  
*D:* Datenflußplan *m*  
*R:* блок-схема потока данных
- check** *n* [E425]  
*D:* Prüfung *f*, Kontrolle *f*  
*R:* контроль *m*, проверка *m*
- check bit** [E412, E441, E449]  
*D:* Prüf-Bit *n*, Kontrollbit *n*  
*R:* проверочный бит, контрольный бит
- Check Channel** (instruction) [E425]  
*D:* Prüfen Kanal (Befehl)  
*R:* контроль канала (команда)
- check character** [E412, E441, E449]  
*D:* Prüfzeichen *n*, Kontrollzeichen *n*  
*R:* контрольный знак

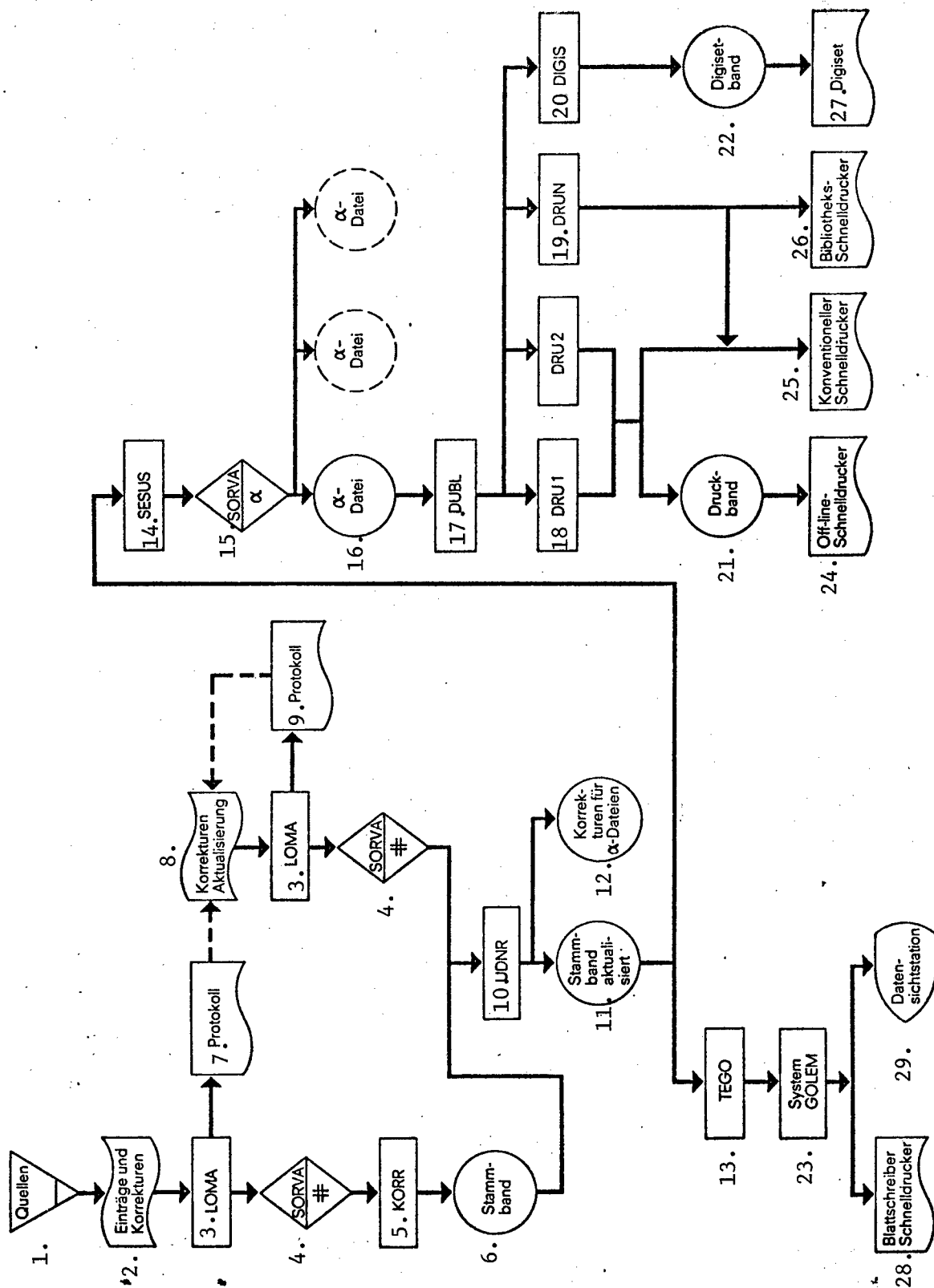
Figure 11. Trilingual Dictionary  
 (Russian in Cyrillic script - DIGISET).

<sup>1</sup> Russian here in the ISO transliteration.



# TEAM Program System Data Flow Chart

## APPENDIX



[Key to Appendix Chart]:

1. Sources; 2. Entries and corrections; 3. LOMA [perforated tape to magnetic tape program]; 4. SORVA; 6. Main tape; 7. Voucher form; 8. Correcting, bringing up to date; 9. Voucher form; 10. UDNR [word store updating program]; 11. Main tape brought up to date; 12. Corrections for  $\alpha$ -data files; 13. TEGO [interface program for remote interrogation]; 14. SESUS [synonym generation program]; 15. SORVA/ $\alpha$  [unknown]; 16.  $\alpha$ -data files; 17. DUBL ["doublet" elimination program]; 18. DRU 1 [printout routine 1]; 19. DRUN [printout routine variant for up to five languages]; 20. DIGIS [data media generator for photo-composer control]; 21. Printout tape; 22. DIGISET tape; 23. GOLEM [a documentation system]; 24. Off-line high-speed printer; 25. Conventional high-speed printer; 26. Library high-speed printer; 27. DIGISET composer; 28. Page printer, high speed printer; 29. Data viewing terminal.

BIBLIOGRAPHY

1. Huelck K., Klugmann D., Peetz G., "GOLEM - ein allgemein anwendbares Verfahren fuer die Dokumentation und das Wiederauffinden von Informationen" ["GOLEM - A Generally Applicable Procedure for the Documentation and Retrieval of Information"], Siemens-Schriftenreihe data praxis [Siemens Data Practice Publications Series], 1967, 12 pp.
2. Pattay W. von, "Fernschreiber 106: eine Lochstreifenschreibmaschine mit erhoehtem Zeichenvorrat fuer Bibliotheken" ["The 106 Teletype: A Perforated Tape Typewriter with an Increased Stock of Characters for Libraries"], NACHRICHTEN FUER DOKUMENTATION, 18, 1967, No 2, pp 72-75.
3. Hahn A., "Automatische Bibliotheksverwaltung" ["Automated Library Management"], DAS RATIONELLE BUERO [THE EFFICIENT OFFICE], 19, 1968, No 12, pp 9-15.
4. Tanke E., "Das aktuelle Woerterbuch aus der Datenbank" ["The Up-to-Date Dictionary from the Data Bank"], DEUTSCHER DRUCKER [GERMAN PRINTER], 1971, No 38.

8225

CS0:8320/0615

## LEXICOGRAPHY WITH TEAM -- AUTOMATIC DICTIONARY COMPOSITION

Munich SONDERDRUCK AUS 'DATA REPORT' in German No 9, 1974 pp 9-13

[Article by Joachim Schulz: "Lexikographie mit TEAM. Automatischer Satz von Woerterbuechern"]

[Text] The TEAM<sup>1</sup> (Terminology Recording and Evaluation Method) was developed in the language service of the Siemens company in Munich as an aid for in-house application, where the central feature of the system is a multilingual dictionary stored in a data processing system. This contains technical and scientific expressions in the most important European languages, where in addition to the individual designations, including their synonyms, yet additional information is stored, whether it be of a grammatical or topical nature, such as definitions, source citations, etc. The totality of this information on one concept forms a so-called entry or word location.

The electronic dictionary, which at the present time consists of a few hundred thousand such entries (and is designed for a few million) can be made accessible to the user in different ways: on one hand, through direct interrogation, for example, via a data viewing terminal, and on the other hand, through indirect means, taking the approach of putting out printed lists, glossaries or dictionaries. Since in light of the quite rapid development of science and engineering, the value of a technical dictionary is not the least of all in how up to date it is, electronic data processing offers itself as a fast and reliable aid in the production of dictionaries. Some of the problems arising in this case, as well as the possibilities for solution which the TEAM system offers, are described in the following by Joachim Schulz of the language service of Siemens AG [Inc.], Munich.

### The DIGISET Photo-Composer

First a brief look at the technical prerequisites (Figure 1). While the programs for storage, correction and processing of the terminology data run on a Siemens 4004/35 data processing system (or a larger one) with a perforated

---

<sup>1</sup> This project is supported by the Federal Ministry for Research and Technology.

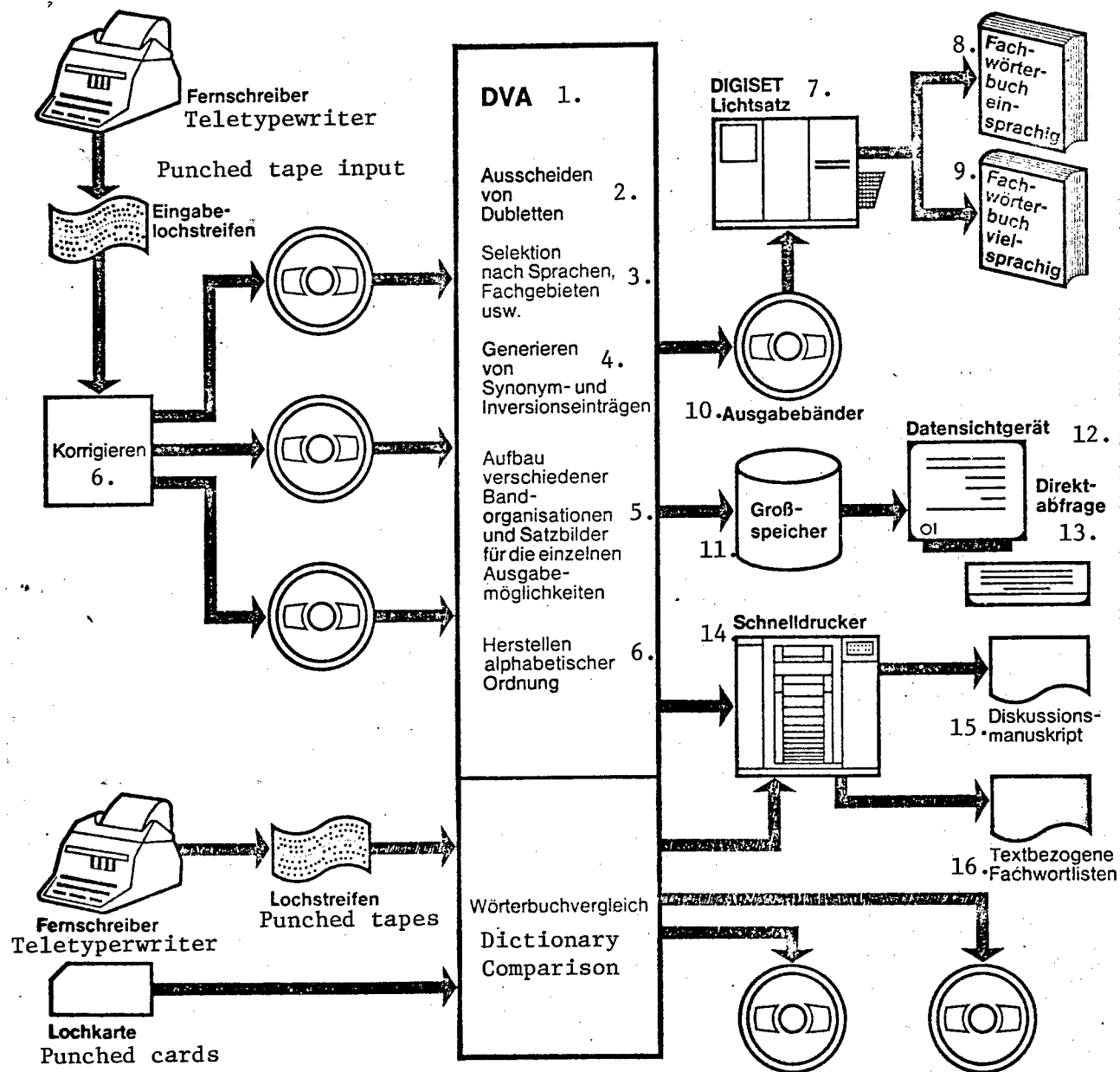


Figure 1. The information stored in the terminology data bank can be made useful in various ways: Putting out dictionaries, direct interrogation in dialog operation, and printing out selected or textually referenced technical word lists in accordance with various criteria.

Key: 1. Data processing system; 2. Exclusion of 'doublets';  
3. Selection according to languages, technical fields, etc.;

[Key to Figure 1 Continued]:

4. Generation of synonym and inversion entries; 5. Structuring various volume organizations and composition patterns for the individual output possibilities; 6. Alphabetizing; 7. DIGISET photo-composition system; 8. Single language technical dictionaries; 9. Multilingual technical dictionaries; 10. Output tapes; 11. Large memory; 12. Data viewing terminal; 13. Direct interrogation; 14. High speed printer; 15. Discussion manuscript; 16. Textually referenced technical word lists.

tape reader, high-speed printer, punched card reader and five magnetic tape units, used for the data output (of the dictionary) is a photo-composing system, which is at the present time probably the most modern aid for automatic composition. The composing program available in the TEAM system is designed for the Digiset<sup>R</sup> 50 T1 system of the Dr.-Ing. Rudolf Hell GmbH Company, Kiel. (At this time, work is underway on an additional program, which, in particular, takes into account the capabilities of the 40 T1.) Two types of input are to be differentiated in these systems: The typographical data and the textual data.

The typographical data, i.e. all information concerning the typography to be used, are first written into the central store of the system. From there, the individual characters of this typography are called up by the textual data for projection on the screen of the cathode ray tube. The total core memory of the photo-composing system (Model D) is to be subdivided into three or four areas, each of which can record one type of typography.

The textual data encompass the characters which are themselves to be used in the composition: Thus, the letters, numerals, composing and special characters, which, as already mentioned, call up the corresponding image pattern from the store, and on the other hand, the control instructions for the positioning and modification of these characters. The letters and characters of a text are displayed one after the other on the screen of a cathode ray tube, and project it from there onto a film (or onto photopaper). A character can be positioned in the vertical and horizontal positions. A modification is possible with respect to size, width and setting (straight or cursive).

#### Preliminary Operations

It is obvious that the actual composition work, i.e. the preparation of the data and the compilation of the text pages, has to take place prior to the technical execution of the composition just described. Three large stages are to be distinguished in the process of the automated preliminary work:

- The selection of a set partial quantity out of the total stock of stored word locations;
- The sorting and compilation of the data in forms customary for dictionaries, and,

-- The production of a printing media, i.e. a film with the completely composed and made-up pages of the dictionary.

### The Selection

Technical dictionaries generally limit themselves to the terminology of set, rigidly circumscribed technical fields (Figure 2). A decisive criterion in the selection of expressions for a particular dictionary is, for this reason, the subject area codes supplementing all entries. Several such codes can be taken into account, i.e. one selects all concepts which belong in at least one of the areas concerned, or only those which are used simultaneously in all the desired areas.

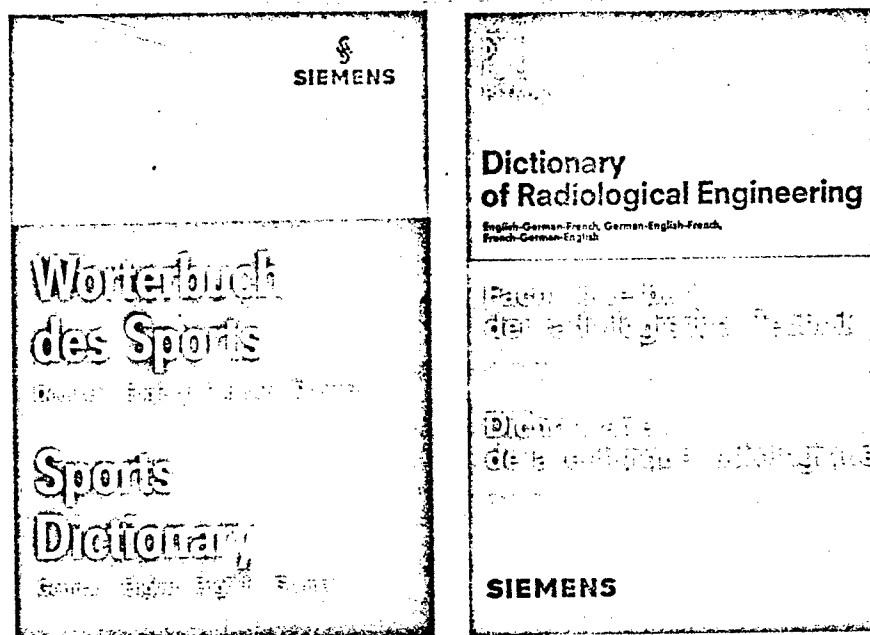


Figure 2. Two examples of technical dictionaries. Others are cited in the bibliography under [4] and [5].

Since also stored in the computer dictionary in addition to the unobjectionably explained and employed expressions are temporary working concepts, a further selection can be made according to the quality of the entries. It is frequently the case that in the source language there is in fact a precise expression available, however, known in the target language is only a helpful expression, a paraphrase. This can naturally be an aid to the translator, but for its part should not appear itself in the source language as a search concept. This problem too is solved automatically through the program based on a certain formal characterization.

Finally, a dictionary will only encompass a certain number of languages. Thus, a selection is necessary which supplies those entries from the total stock

which contain the information in the desired languages. Tied to this is the possibility of a broader selection, and in fact, within the word locations themselves: it can be determined through program parameters just what additional information should be fed out or suppressed for the particular expressions (possibly synonyms, source citations, definitions, etc.).

### The Sorting

Of course, all languages of the stored entries can be selected as the output language and the data sorted according to these languages. The basis for sorting the expressions is a so-called sorting concept, which is set up through the program for the particular output language. It makes automatic and exact sorting possible according to the rules of the different languages and alphabets. Thus, quite generally upper case letters are sorted as lower case letters; in German, the umlauts ä, ö and ü are classified as the analogous basic letters, a, o and u, the accents are not observed in French, and in Spanish, "ñ" and "ll" receive one individual position value. Naturally, the Russian expressions are arranged in accordance with the rules of the Russian cyrillic alphabet, although they are recorded and stored in transliterated form, i.e. using Latin letters and accents.

Above and beyond this, the sorting can be carried out so that expressions which consist of several words ("logic circuit"), are alphabetized going straight through, without considering the space between the words, or in a manner though such that this intermediate space is quite probably taken into account, so that the expressions are first compiled which contain the same word (as the "key word") in the first position, and then those in which it is part of a larger (compound) word ("logic unit" before "logical").

119	Zeichenträger
<i>R</i> : память с пословной организацией <b>wortorientiert</b> <i>adj</i> [E42] <i>E</i> : word-oriented <i>adj</i> <i>R</i> : ориентированный на слово <b>Worttaktzeit</b> <i>f</i> [E422] <i>E</i> : word time, word period <i>R</i> : длительность слова, время обработки слова <b>Wortzeit</b> <i>f</i> [E422] <i>E</i> : word time, word period <i>R</i> : длительность слова, время обработки слова <b>Wurzel</b> <i>f</i> [E11, E4201] <i>E</i> : base <i>n</i> , radix <i>n</i> <i>R</i> : базис <i>f</i> , основание <i>n</i>	<i>R</i> : изображение знаков, представление знаков <b>Zeichendichte</b> <i>f</i> [E4114, E4115, E4116, E4117] <i>E</i> : recording density, character density <i>R</i> : плотность записи <b>Zeichendrucker</b> <i>m</i> [E4113] <i>E</i> : character-at-a-time printer, character printer <i>R</i> : печатающее устройство знаков, буквопечатающее устройство <b>Zeichenerkennung, automatische</b> [E4281] <i>E</i> : character recognition <i>R</i> : распознавание знаков, опознавание знаков <b>Zeichenkonstante</b> <i>f</i> [E425] <i>E</i> : character constant <i>R</i> : константа знаков <b>Zeichenkonzentrator</b> <i>m</i> [E4114] <i>E</i> : pack/unpack facility
<b>Z</b>	
<b>Zahlendarstellung</b> <i>f</i> [E11, E4201]	

Figure 3.

Sample of a trilingual German-English-Russian Data Processing Dictionary presently in preparation. The language elements are arranged vertically and provided with locating aids.

## Synonyms

Two additional problems arise in connection with data sorting. The first question concerns synonyms, which as already noted, are incorporated together with the basic designations for a concept in a main entry. While in the target language or languages of a dictionary all existing synonyms can be given one after the other without further ado, synonyms possibly present in the source language must appear at quite different places (in accordance with the way they are written) in the dictionary. There they should either have the total information of the main entry with them or also only supply a reference to it, so that the user can find the desired information there. In order to satisfy this requirement, prior to the actual sorting by the program, all the possible synonym entries or references are generated, which are associated with the main entries for the concepts in the source language of a dictionary.

check	28
<p><b>check</b> <i>v</i> kontrollieren <i>v</i>, nachsehen  <i>v</i>; to - a rope (YACHTG) ein Ende  schricken; to - way (YACHTG) die  Fahrt verringern; to keep in - in  Schach halten  <b>cheek</b> <i>n</i> (EQUEST) Knebel <i>m</i>; curb-  bit with curved -s and port  (EQUEST) gewöhnliche Kandare  mit gebogenen Müllern und  Zungentrenner; egg-bunt snaffle  with -s (EQUEST)  Olivenkopftrense mit Knebeln;  snaffle - (EQUEST)  Trensenknebel <i>m</i>; snaffle with -s  (EQUEST) Knebeltrense <i>f</i>  <b>cheek-piece</b> (of the bridle) <i>n</i>  (EQUEST) Backenstück (des  Zaums) <i>n</i>  <b>cheek-strap</b> <i>n</i> (EQUEST)</p>	<p>- timekeeper (SWIMG, TRACKF)  Zeitnehmerobmann <i>m</i>  <b>chin</b> <i>n</i> (MED) Kinn <i>n</i>; - (of the  horse) <i>n</i> (EQUEST) Kinn (des  Pferdes) <i>n</i>  <b>chine-type</b> <i>adj</i> (YACHTG)  knickspantig <i>adj</i>  <b>choice of baskets</b> (BBALL)  Korbwahl <i>f</i>; - of courts (VBALL)  Wahl der Spielfeldlinie; - of  ends (VBALL) Spielfeldwand <i>f</i>;  Seitenwahl <i>f</i>; - of exercise  (GYMN) Übungswahl <i>f</i>; - of  service (VBALL) Wahl der  Aufgabe; option of - of ends  Recht der Seitenwahl  <b>chokelock</b> <i>n</i> (Shime Waza) (JUDO)  Würgegriff <i>m</i> (Shime Waza);  <b>naked</b> - (Hadaka Jime) (JUDO)</p>

Figure 4. Sample page from the Sports Dictionary. The language lines are arranged horizontally. When repeated, key words are replaced by a tilde.

## Inversions

A second, similar problem comes up in the case of expressions which consist of not just one word. Multiple word designations (compound expressions) occur especially frequently in technical language. These multiple word expressions are as a rule recorded and stored taking into account the natural word sequence of their components (for example, "bistabile Kippschaltung" ["bistable sweep circuit"]: the adjective comes before the noun). In translation, thus in the target language of the printed dictionary, this word positioning is also maintained. However, in the source language the desired information (the translation of this compound expression) should not be found under only



the first word (here the adjective), but also with the possibly subsequent words (here the noun), insofar as these have sufficient weight of their own. These so-called key words are specially marked early in the data recording, i.e. provided with a control character, by means of which the so-called inversion or reverse entries are generated through the program. These inversions are independent dictionary entries in which the key word is placed in the first position of the designation, and the remainder of the expression added on after a comma. (Thus, the term "bistabile Kippschaltung" is to be found as an additional inversion entry, "Kippschaltung, bistabile", and is consequently found both under "b" and "k".)

The marked key words can, with regard to an appropriate sorting and arrangement in the dictionary, also be preferred if they occur as part of a compound or in a larger expression in inflected form: for example, "Verknuepfung, NAND-" ["Gate, NAND-"] to "NAND-Verknuepfung" or "Kamera, mit zwei ~s aufnehmen" ["Camera, photograph with two ~s"] to "Mit zwei Kameras aufnehmen". This last expression will appear in the German section of the dictionary only in the inverted form shown here, since it does not make sense to categorize under the initial word "mit" ["with"].

#### Automatic Composing

Remaining as the third and last stage in the automatic production of a dictionary is the generation of the print medium by means of the electronic Digiset photo-composition system mentioned above. The operation of this system is advantageously controlled in off-line operation by a magnetic tape, on which the text data (the dictionary entries) are found along with the requisite typographical instructions. A special composition program (to produce this magnetic tape) must therefore convert the previously selected and sorted word locations from the TEAM format (EBCD code) to the form required by Digiset (primary addresses) and simultaneously provide them with all the requisite control instructions for the composition of complete dictionary pages. The original store format of the data permits an adequate differentiation of the individual information components at each word location, so that these can be placed in any order and set in any type (for example, in different types of writing).

The typographical structure of the dictionary is not established when the data is recorded, and that is to be emphasized here once again; not until this point in the processing are the corresponding control instructions for the Digiset generated through the program. The selection and sequencing of the partial information to be set up (the designations in the various languages, as well as individual supplemental information entries, are likewise not determined through program parameters until now).

#### Stock of Characters

Because of the large stock of characters of the Digiset system, the text can be put out in true orthography. (A prerequisite for this is naturally a corresponding recording and storage of the data, a condition which is met in the

TEAM system.) Besides the upper and lower case letters in an alphabet (including the umlauts and the "ß"), as well as the punctuation marks and numbers, there is an entire series of special characters available, including various forms of quotation for various languages, the ligatures used in French, œ, Œ, etc. With a greater expansion stage of the main memory of the Digiset 50 T1, more than the three scripts mentioned at the outset can be made available simultaneously.

### Accents

Accents and cedillas pose a problem in the composition preparation insofar as they (so as not to expand the alphabets for the various European languages necessarily) are fed in separately from the basic letters and must also be set up in this way (so-called fleeting accents). During data recording, they are written out as on a standard typewriter, ahead of the corresponding letters (keyed).

For exponents and subscripts in chemical formulas, as well as in mathematical and physical expressions, likewise no special typographical characters are employed: the corresponding characters (numerals or letters) are composed first by the program control instructions, which see that the characters are correspondingly reduced in size, as well as shifted up or down.

### Cyrillic Letters

An additional problem comes up when composing cyrillic texts (Figure 3). Since Russian words, as has already been mentioned, are recorded and stored in Latin transcription, the characters of the cyrillic alphabet must be again back-transliterated for the output following the input transliteration [3]. The scheme recommended by the ISO (recommendation 9) is used for the transliteration, and is essentially based on the Czech alphabet.

### Line Construction

The dictionary text is processed word by word in constructing the lines, where the thicknesses of the individual letters are added, and a check is made to see whether the specific lines contain yet enough space for the next word. If so, the word is placed in the line; if not, a new line is started. A word is broken up only with a hyphen (or a slash line). Automatic syllable separation is not undertaken in the program system, since corresponding programs are not yet available for all of the languages used. Moreover, composition without justification in the case of dictionaries, which really offer no continuous text, proves to be not at all disruptive.

### Word Locations

A word location always begins with the expression of the source language which is set in medium-faced print. Additional explanations in parentheses or references appear in thinner typescript after this. Specifications of the parts of speech are set in cursive type. Since data on the technical field for each

---

## The Computer Looks It Up

Besides putting out up to date dictionaries, the TEAM system also offers the translator direct access to the stored terminological data.

Since direct access with the dialogue capability is not always required and in light of the expense is not always justified, above and beyond this one more especially tailored batch interrogation procedure was developed for translation work in a large language service.

This procedure takes over from the translator the often tedious and time consuming work of looking up unknown words. For this purpose, the translator only needs to underline the technical expressions in the text to be translated which are unfamiliar or unknown to him. These expressions are then transferred to punched cards or perforated tapes, fed into the computer, and compared there with the stored dictionary. In case compound expressions are not found in the dictionary, their component parts can be treated as a type of supplemental question and be "looked up". Finally, all questions, together with the answers found are put out on a list one after the other in the sequence of their occurrence in the text.

Such "textually referenced technical vocabulary lists" not only lead to a productivity increase of over 50% as was demonstrated as early as 1965 in the then Translation Service of the Bundeswehr, it also assures precise and uniform terminology in the case of extensive translation projects which are distributed among several coworkers.

---

term are extraordinarily important for the translator (they refer to the particular technical applications area and assist in resolving questions of homonyms or words with several meanings), these are generally fed out along with them. A corresponding data key code, or even several, are as a rule put in square brackets following the term in the source language.

The arrangement of the target languages, which are set in increasingly lighter type, is possible two forms: in a vertical configuration (Figure 3), the expressions in the various languages are set up specifically on a new line so that the individual language elements within a word location appear one under the other. A short language key is set up first in this case as a locating aid, i.e. an "E:", "F:", etc. for "English", "French", and the like. The horizontal form (Figure 4), on the other hand, places all the information of a word location in a continuous sequence. Where needed, the locating aids cited above can also be incorporated so that it is possible to put out any number of languages, as in the case of a vertical structure. However, without locating aids, the horizontal form appears to be especially suited to a two language dictionary.

## Paragraph Formation

In a bilingual dictionary with a horizontal structure there is the additional possibility of paragraph formation: word locations, which contain the same key word in the source language, can be combined in groups. This combination is made possible by the already mentioned generation of inversion entries and a corresponding sorting of the data. Within the groups or paragraphs, the key word is replaced by a tilde where it is repeated.

## Page Structure

If a dictionary page is to be composed of two columns, where each of them is to be read from top to bottom, each page must be completely structured in the core memory of the composing computer before it can be put out line by line (for two half-lines each). Furthermore, each page must be provided with its pagination, and where needed, with a current column heading.

For locating assistance, the individual columns in which the word locations are arranged one under the other, or even the entire page, can be interrupted by a set number of blank lines, when the initial letter changes. This letter is inserted, likewise automatically, in a higher type face in the blank space.

## Corrections

Naturally the entries in an electronic dictionary can still contain errors. Even if the computer or the program is not in error, all typographical errors which are made by people during the data input are again visible with the output. Thus, proofreading is also not left out in the automatic composition processing. Normally, a discussion manuscript is printed out on paper prior to the final composition on film. More extensive corrections can be made then so that the source data stored in the data processing system, which are really to be corrected in any case, are emended and finally the processing is run through again up to the point of composition. In the case of small errors, a so-called secondary correction is possible in which the errors are eliminated through hand set-up of the film. In any case, the output (even subsequently) of individual pages of the dictionary is possible on the Digiset, since the text is broken down by page using special section markers for the magnetic tape input unit of the Digiset.

## Experience

A data processing dictionary was produced by the language service for the first time in the fall of 1970 by means of the procedure described here. Siemens AG was the editor and publisher [4]. In the meantime, in part in cooperation with other publishing houses and authors outside of our company, an entire series of additional dictionaries have followed, a few of which are pictured (p. 11) [5]. It was possible to work numerous changes and improvements into the program system based on the experience gained in this case and the particular diverse requirements.

The primary advantage of the procedure should be noted once more by citing a few production times: about 15 minutes were required for the selection of around 10,000 entries from an overall stock of more than 200,000 and the simultaneous generation of synonym and inversion entries, about 8 minutes were needed for the sorting and finally about 6 minutes for the composer program. Thus, in approximately a half an hour (absolute computer time, i.e. without down time and naturally without correction operations) all the expensive selection, sorting and composition operations were taken care of. The film composition itself then lasted for only about 45 minutes.

Added to this is the fact that the electronic dictionary, which is the basis for this work, can be kept continually "up to date", a requirement which is today indispensable in the fields of technical languages, and also offers the simple publication of corrected, new editions. Thus, electronic processing makes it possible to also publish conventional dictionaries, which are not already obsolete the moment they appear.

#### BIBLIOGRAPHY

1. Brinkmann K.-H., Tanke E., "TEAM - ein Programmsystem fuer die Loesung terminologischer und lexikographischer Aufgaben" ["TEAM - A Program System for the Solution of Terminological and Lexicographical Problems"], Siemens Publications Series Data Practice, (Order No. D14/4059).
2. Brinkmann K.-H., Schulz J., Tanke E., "Das Woerterbuch aus der Maschine" ["The Dictionary from the Computer"], DATA REPORT 4 (1969), No. 4, pp 9-15.
3. Schulz J., "Reversible Transliteration kyrillischer Buchstaben" ["Reversible Transliteration of Cyrillic Letters"], NACHR. f. DOK., 24 (1973), pp 239-241.
4. Woerterbuch der Datenverarbeitung/Data Processing Dictionary, Vols 1 and 2, Berlin, Munich, Siemens AG, 1970
5. Feldbausch F.K., "Bankwoerterbuch [Banking Dictionary] (German-English/English-German), Munich, Moderne Industrie Publishers, 1972.

Dictionary of Radio and Television Terms/Fachwoerterbuch Hoerfunk und Fernsehen (English-German/German-English), published by ARD in conjunction with BBC and ZDF, Berlin and Munich, Siemens AG, London, Pitman Publishing, 1972.

Woerterbuch des Sports [Sports Dictionary], Vols 1 and 2, (German-English/English-German and German-French/French-German), Official Dictionary of interpreters and translators at the XX Olympic Games, Munich, 1972, Berlin, Munich, Siemens AG, 1972.

8225  
CSO:8320/0615

COMPUTER ASSISTED TRANSLATION PROCEDURE FOR QUERYING A MULTILINGUAL TECHNICAL  
DICTIONARY STORED ON MAGNETIC TAPE

Munich LEBENDE SPRACHEN in German Vol 20 No 4, 1975 pp 1-4

[Article by J. Schulz: "The Computer Helps the Translator. A Procedure for  
Querying a Multilingual Technical Dictionary Stored on Magnetic Tape"]

[Text] Previous efforts to carry through translations by means of computers  
have not led to results which permit replacing men by computers, or even al-  
low us to anticipate this in the foreseeable future. The continually increas-  
ing flow of scientific, technical and commercial information, which is to be  
continually translated, makes it appear necessary and meaningful though to  
employ the computer as an aid for the translator.

A terminological data bank was set up in the language service of the Siemens  
company for this purpose, and the TEAM<sup>1</sup> program system was developed, which  
makes possible the management and processing of the information stored in it  
to serve the translator [1, 2, 3]<sup>2</sup>.

The primary functions which this data bank serves are, on one hand, the cen-  
tral storage of the terminology of the technical fields represented at Siemens,  
as well as keeping it up to date (including related and applications areas),  
and on the other hand, is an effective information service which makes the  
stored information available to the translators in suitable form. Two aspects  
are to be considered in this case: On one hand, there should be assurance that  
even with large numbers of coworkers, the uniformity of the terminology is  
maintained, so that each individual translator does not come up with different  
technical expressions from different sources for the same subject. Above and  
beyond this, the translator should be relieved of time consuming routine work  
through at least partial automation of the translation process, especially  
in looking up precise technical expressions in the target language. In other  
words, the specifically desired terminology should be supplied from the data  
bank automatically or semi-automatically for the text to be translated.

The information capabilities which the TEAM system offers extend from the  
automatic printing of technical dictionaries, which are composed by means of  
a DIGiset electronic photocomposing system [4], to individual interrogation

in dialogue via a data viewing terminal. For the nonparticipant, this latter form is often the most impressive, though one can see that it would neither always justify the expense, nor, considering the subject material, even always be desirable or necessary.

### Technical Glossaries

Whenever greater partial quantities of the stored terminology stock are required, thus, when greater numbers of questions are not to be answered individually and directly, a sequential, so-called batch processing or 'Stapel-Verarbeitung' of the information stored on magnetic tape makes sense. In order to localize this interrogation capability somewhat more precisely: the desired body of terminology is not defined in terms of one or more technical fields (as for example, in the printing of selected technical glossaries), but rather through a specific body of individual questions which are directed to the dictionary. These questions can arise during terminological and lexicographical work, when for example, a check is to be made as to whether certain expressions are already stored, or which equivalents are already recorded for them in another language. The questions can also be asked though (and this will be the most important applications case) by the translator as preliminary work for the translation of a particular text. The results of such queries are so-called textually referenced technical glossaries, the successful use of which was studied and practiced for the first time in the former Translation Service of the Bundeswehr [5]. With the textually referenced interrogation in the TEAM system, a distinction can be drawn even between an alphabetical and a textual or reading-synchronized output in accordance with the arrangement of the answers on the high speed printer list supplied as the result. Both contain, in different configurations, the desired technical words (the "questions") and along with them the corresponding dictionary entries as answers from the data bank.

Before covering the interrogation possibilities themselves, the terminological data stored in the TEAM system is to be described briefly. In conclusion, a few more questions of the technical and program engineering realization will then be treated, insofar as they are of importance for the user of the procedure.

### Dictionary in the Computer

It can be quite generally said that in the case of the TEAM system terminological data bank, we are dealing with a multilingual technical dictionary which at the present time encompasses a few hundred thousand concepts, where the dictionary in its simplest form is stored on magnetic tape. The most important information contained on it is naturally the technical words themselves, more precisely, the simple or compound designations for technical concepts, which are recorded in up to eight different languages (the most important Western European languages and Russian). All languages stand on an equal footing with respect to each other, and the equivalents in the different languages, including any number of synonyms, can be followed simultaneously, or later at any point in time. As a rule, the terms are accompanied by a number

of supplemental information entries: for example, simple grammatical data (such as the part of speech), source references (where the term is to be found), a definition of the concept, one or more technical area codes (i.e. a precise assigning of the concept to set technical fields), and various administrative data. All of these partial information entries, which are referenced to a concept, which do not always have to be present in their full number, and are only retrieved from the store in their true, i.e. variable length, comprise a so-called entry, which represents the information unit in the system.

For batch interrogation (just as for lexicographical and other batch operations), an alphabetically sorted version of the dictionary is produced and maintained for each processed source language. Generated on these tapes in the source language concerned for the stored synonyms are their own complete entries, which appear at their own place in the alphabet. The same applies to the so-called inversion entries for multiple word designations. While these expressions are first of all recorded and stored in their natural word sequence (in German, for example, the adjective before the noun, as in "bistabile Kippschaltung"), the desired information can be found by means of the inversion entries not just under the first word, but also under important subsequent words (for example, "Kippschaltung, bistabile"). Abbreviations are treated in a similar fashion with their corresponding written-out forms (for example, "EDV, elektronische Datenverarbeitung", along with "elektronische Datenverarbeitung, EDV").

The questions which can be answered by the system are then technical words or technical expressions compounded from several words in a particular language, for which the corresponding equivalents can be supplied in another language as the answer.

### Interrogation Possibilities

Before the system can process the questions and find the answers, some preliminary operations are to be carried out. In the case of the textually referenced query, one proceeds in the following fashion: all unknown expressions are underlined in the text to be translated and they are thus isolated from the casual textual relationship. At the same time, one places them in standard form for the dictionary by striking the inflected endings out. In the simplest case, there is a possible question of one word in the grammatical, basic form (nominative singular, infinitive, etc.). The same applies to the particularly frequent multiple word designations in the technical languages (for example, adjective--noun combinations, compounds in English), and in this case an entire group of words is underlined, and where necessary placed in dictionary standard form by underlining the endings (for example, ... Technik der gedruckten Schaltungen ...).

### Key-Word Questions

In order to set the system up in a more flexible fashion above and beyond these simple query possibilities, some semi-automatic aids for "looking things up" are additionally present in the computer dictionary. So-called key word



questions can be asked as a special form. The answers to these questions (of the computer, for example, to be especially marked by a "\*" placed in front) are all entries in the dictionary which contain the "key word" concerned. Strictly speaking, the issue in this case is not one of individual words, but one of arbitrary letter sequences (for example, \*Kipp . . . or \*Schalt . . .). This query possibility takes advantage of the existence of the already mentioned inversion entries in the dictionary, in which important word components in multiple word expressions (or in compounds) are placed at the beginning, so that these compounds can be found in a simple comparison.

#### Supplemental Questions

A further interrogation possibility is especially helpful in searching for multiple word expressions. Specifically, it is not always clear from the outset which words in the text should be incorporated into a question, i.e. whether they correspond to a compound technical expression in the dictionary. On the other hand, it can be of considerable assistance, when a sought expression is not found at all, to at least obtain the translation of certain component parts. For this purpose, the so-called sequential or partial questions can be asked in the form of supplemental questions, which, however, do not have to be additionally written out and fed into the machine. It is enough to mark the parts concerned (by a control character, perhaps a "+") in the original total printout (for example, rewinding to the + tape start mark). As noted above, these supplemental questions are answered (i.e. translated) only if the encompassing expression is not to be found in the dictionary. On one hand it is possible to incorporate several individual words of a long expression into a sequential question, and on the other, partial components can also be extracted from a compound and disruptive inflection endings for a partial expression eliminated in the interrogation (for example, tape + start mark, + steckbar/e + Schaltplatte [+ plug-in + circuit board]). The marking characters, just as other special characters, are passed over during dictionary comparison. Furthermore, each partial question can itself be again marked as a key word question and handled correspondingly.

In addition to the preparation of the questions described here, the translator has to set some general data for their processing. In particular, he must communicate to the system what the source language is and what information is required from the dictionary, i.e. whether the dictionary entries with all of the partial information contained in them is to be fed out, or whether only one or more target languages with specific supplemental information (definition, source, etc.) are desired.

#### Machine Processing

The requisite preliminary work of the translator ends at this point. Before the actual machine processing begins, the prepared questions, along with the code data mentioned above, must be put in a form which the machine can read. For the most part, punched cards or perforated tapes are used for the input. In this case, each punched card contains only one question (including the

markings for the subsequent questions); in making perforated tapes by means of a teletype, following each question, the line space is keyed.

The processing in the computer, generally speaking, runs its course in the following fashion: the questions are read into the working store of the system and possible subsequent questions are generated simultaneously. All questions are sorted in the store, then the dictionary magnetic tape is read and the questions are compared with the dictionary entries. In case of correspondence, the entries are extracted and stored on an intermediate magnetic disc. In conclusion, the desired sequence is again established, and the questions, as well as the answers are fed out through a high speed printer.

### Question Quantity

When reading in the punched cards or perforated tapes, the question arises as to how large the number of questions can be ("the batch of questions"), which is to be handled at one time in the system. Since the store capacity corresponding to the particular quantity of questions can be assigned to the interrogation program up to the full capacity of the particular computer system (for example, 40 to 250 Kbytes), and as a rule of thumb, about 5 Kbytes are required for 100 questions, a maximum of 5,000 questions are to be processed in one program run at the present time. If this number is exceeded, the questions are automatically subdivided into batch sections, and these are processed sequentially. It is advantageous to put all question batches (of different translators also) together with the same source language, independently of the target languages required, so that these questions can be answered in one dictionary run-through.

### Stock of Characters

In storing the questions, one simultaneously sets up a particular standard form for them in order to compensate for various possible written forms and provide for an effective comparison with the similarly handled dictionary entries. In this case, several problems are to be considered. For example, while the punched cards have at their disposal only a limited stock of characters (capital letters, numbers, and a few punctuation marks), there are also available for the recording of the questions on the teletypes used at the present time lower case letters, umlauts and numerous diacritical marks. The dictionary information itself is stored in the correct orthography, i.e. with the complete stock of characters applicable to the particular language. (The Russian cyrillic alphabet is transliterated using latin letters and diacritical marks in accordance with the ISO recommendation [6]). With questions in simplified form fed in via punched cards, the letter combinations of AE, OE and UE in German are replaced by the corresponding umlauts (if they are not to remain separate, something which is checked against a special list). By forming the standard form, upper case and lower case letters are then made equal, "ß" replaced by "ss", and all accents are eliminated. Above and beyond this, all hyphens and the gaps between multiple word expressions are suppressed. This is particularly helpful in English, where numerous compounds can be written with or without a hyphen. The dictionary search, which is really

based on a character by character comparison of questions and possible answers, can in this way be successful even with different forms of writing. (Of course, other writing variants, as for example between American and British English, are not to be disposed of in this fashion.) The standard form makes an unambiguous alphabetical sorting of the questions possible in the system, which corresponds to the sorting in the dictionary. (Strictly speaking, the questions themselves are not sorted in the memory, rather only their addresses, which is accomplished much more quickly and simply, and above and beyond this, doesn't require any back-sorting of the questions and their answers prior to the synchronous text output.)

### The Comparison Length

A further problem which arises in sorting, as well as in the interrogation, is the sorting or comparison length. While the questions, just as the dictionary entries, can be of variable length, it is expedient to work with a fixed length for the subsequent processing, where this length can be set arbitrarily though. A compromise length of 30 characters has proven to be good for sorting dictionaries. For sorting the questions in the store and for the comparison, a length of 20 characters was selected. The most important reason for this is the capability of saving space and being able to simultaneously process several questions in the store. On the other hand, a study of the existing dictionary entries has shown that the number of entries which coincide with others as regards this limited comparison length cannot be substantially reduced any further if the length is increased from 20 up to 25 or 30. The relationship of the maximum number of such "doublets" (and thereby of the possible multiple answers to a question) to the comparison length shows this same tendency: the sharpest decrease in this figure is achieved with an increase of up to 20 characters. This limitation of the standard form for the dictionary query has the consequence that considered as an answer is a dictionary entry in which the corresponding technical expression does not differ from the question in the first 20 letters.

### Answers from the Dictionary

A sequential processing of the dictionary is possible based on the alphabetical sorting of the questions, i.e. the relevant magnetic tape needs to be run through once in one direction. (A special case: if following a key-word question the same word is queried individually one more time, then the corresponding place in the dictionary can be overlooked, and then one must "page back through", for example, if after all compounds with "bistabil" the same adjective is alone queried once more, possibly by another user.)

When questions appear several times, something which can be the case with longer texts or when there are several simultaneously questioning translators, the dictionary only has to be referred to once. The answers, more accurately, their addresses are then noted for all the identical questions. (The answers themselves, i.e. the dictionary entries in their full length which have been found, are stored on an intermediate magnetic disc.)

In this respect, the question of the possible number of answers comes up. In the case of the so-called key word questions, one expects several answers from the outset, specifically all expressions which begin with the given sequence of letters. Multiple answers are also possible in principle with all other queries: to be considered as answers are all entries which (considering the already mentioned normalization and the fixed comparison length) are in agreement with the question. The maximum number of answers to be fed out is at the present time around 600 (a number which is hardly actually used in a meaningful sense). At the present time, the technical field to which the expression is applicable is not taken into account in selecting the answers. The user himself has to make this choice in light of the supplemental information of the dictionary entry. The same applies for taking into account the part of speech in the case of corresponding homographs.

Supplied as answers are also those entries in the dictionary which do not contain the desired target language. (Gaps in the dictionary are to be determined and handled in this fashion.) Actually, a question, for which only one entry of this type is found in the dictionary as an answer, is considered to be unanswered insofar as in this case possibly marked subsequent questions are taken into consideration, just as in the case of questions for which no corresponding dictionary exists at all.

The result of a query program sequence, as already noted, is a high speed printer list which contains the questions along with the answers or with the annotation "lacking". The questions appear on the left side of the list, numbered sequentially in each translator assignment in the order in which they are fed in, and the corresponding answers appear on the right side (numbered according to the question). Besides this textually synchronous output, an alphabetical ordering is also possible. In this case, questions which are asked many times are listed only once. The lists can be printed out directly on a high speed printer, or following intermediate storage on a magnetic tape or magnetic disc, by which means the actual search sequence is accelerated. Two different types of printers come into play. The printers which have only the conventional stock of characters available (of the punched card), make a conversion, i.e. a simplification in the manner of writing (which is accomplished automatically) necessary. Printout with the expanded stock of characters can be accomplished via so-called library printers.

#### Search Times

The processing times for a batch of questions depend on a multiplicity of factors, as for example, the type of installation (i.e., the internal computer speed and the read rate of the magnetic tape terminals), the level of simultaneous operation in the system, as well as on the type of questions posed (number of key word and follow-up questions). For this reason it is difficult to make a general statement about the response times. Of course, it always makes sense to process as many questions as possible at once in one run-through. The dictionary tape (or tapes) only need to be searched through once in this case, and the percentage of the overall response time which goes to one questions falls off. A few examples will be cited as

points of reference. About 10 minutes are needed for 45 questions, and about 28 minutes plus 18 minutes of printout time for 2,500 questions. The dictionary with about 230,000 entries takes up about two and one half magnetic tape reels in both cases.

#### Dialogue Interrogation

In an effective system of translation aids, the batch interrogation capability described here is certainly not replaced by, but in fact has to be supplemented by a dialogue system with direct access to the stored information. This requires storage of the dictionary on magnetic discs, instead of on magnetic tapes. Running at the present time in test operation in the TEAM system is the first version of an interrogation system developed especially for dictionary search. With it, the translator has the capability of feeding individual questions (which come up along with or following the processing and evaluation of the textually referenced lists) directly through a data viewing terminal into the computer and of receiving the answer directly on the screen. Independently of this, work is going ahead on improving the existing batch interrogation (for example, through automatically taking into account the inflected endings of text words) or replacing this batch interrogation altogether by automatic interrogation. There should thus be made available for texts, which exist in machine readable form, automatic lists of all the technical words and technical expressions contained therein, together with their translations, in order to even further relieve the load on the translator in searching for equivalents in the target language.

#### FOOTNOTES

- <sup>1</sup> Terminology Recording and Evaluation Method
- <sup>2</sup> The work on which this report is based was funded through the Federal Minister for Research and Technology within the framework of the Data Processing Program (Registration No. DV 5,000). The responsibility for contents belongs solely to the author though.

#### BIBLIOGRAPHY

1. Brinkmann K.-H., Schulz J, Tanke E., "Das Woerterbuch aus der Maschine" [The Dictionary from the Computer], DATA REPORT 4, (1969), No. 4, pp 9-15.
2. Tanke E., "Das aktuelle Woerterbuch aus der Datenbank - Ein Beitrag zur Loesung sprachlicher Probleme im internationalen Informationsaustausch" ["The Up-To-Date Dictionary from the Data Bank - A Contribution to the Solution of Language Problems in International Information Exchange"], DEUTSCHER DRUCKER, No. 38, (1971), pp 2-18.

3. Brinkmann K.-H., "Ueberlegungen zum Aufbau und Betrieb von Terminologie-Datenbanken als Voraussetzung der maschinenunterstuetzten Uebersetzung" ["Considerations in the Structuring and Operation of Terminology Data Banks as a Prerequisite for Machine Assisted Translation"], NACHRICHTEN FUER DOKUMENTATION, 25, (1974), No. 3, pp 99-104.
4. Schulz J., "Lexicographie mit TEAM. Automatischer Satz von Woerterbuechern" ["Lexicography with TEAM. The Automatic Composition of Dictionaries"], DATA REPORT 9, (1974), No. 1, pp 9-13.
5. Krollmann F., Schuck H.-J., Winkler U., "Herstellung textbezogener Fachwortlisten mit einem Digitalrechner - ein Verfahren der automatischen Uebersetzungshilfe" ["Producing Textually Referenced Technical Glossaries with a Digital Computer - An Automated Translation Assistance Procedure"], BEITRAEGE ZUR SPRACHKUNDE UND INFORMATIONSVARBEITUNG, 5, (1965), pp 7-31.
6. Schulz J., "Reversible Transliteration kyrillischer Buchstaben" ["Reversible Transliteration of Cyrillic Letters"], NACHRICHTEN FUER DOKUMENTATION 24, 1973, pp 239-241.

8225

CSO:8320/0615

# THE USE OF THE LEXICOGRAPHICAL BRANCH OF A DATA BANK SYSTEM TO PRODUCE A PHRASEOLOGICAL TECHNICAL GLOSSARY

Munich LEBENDE SPRACHEN in German Vol 19 No 3-4, 1974 pp 1-12

[Special reprint of the article by R. Schmidt and O. Vollnhals of Siemens AG: "Der Einsatz des lexikographischen Zweigs eines Datenbanksystems zur Herstellung eines phraseologischen Fachglossars"]

[Text] The following article describes the attempt to solve the information retrieval problems of the language service of the Foreign Office with a data bank system which is already in operation, specifically with the TEAM<sup>1</sup> system developed by the language service of Siemens AG. The TEAM system had already become well known there through several publications. The issue was to investigate whether the transition from a purely terminological data bank with single and multiple word designations to a system which also permitted the processing of technical phraseology is readily possible, since completely different structural conditions really apply to the field of phraseology; considered here are only inflecting endings, inflected irregular verbs, etc. The criteria for retrieving the information (descriptors) must likewise be structured differently. The course and result of the investigation are presented in the following.

## I. Preliminary Remarks

In the course of the enormous growth of the international information exchange, the language service of the Foreign Office, just as other institutions which are engaged in translator and language intermediary activities, sees itself facing a rapid increase in the relevant technical vocabulary. In the case of the technical vocabulary stocks of the language service of the Foreign Office, the issue is, for the by far and away greatest part, not one of simple designations or word equivalents, but one of original text entries from multilingual publications, accords, etc., as well as from other documentation of an official nature, which is to be considered as obligatory in a linguistic sense

---

<sup>1</sup> Terminology Recording and Evaluation Method. The development of the TEAM system is supported by the Federal Ministry for Research and Technology.

both as regards its foreign language equivalence and its specific formulation.

While a technical translator equipped with the requisite subject knowledge is immediately served with the correct terminology for mastering a technical text, here one must absolutely fall back on original text entries. Conventional reference works are thus largely excluded as a translation aid for this special field.

On the other hand, the vocabulary and the associated sources can still hardly be overlooked for the individual, and for this reason must be maintained in some kind of clear and easily accessible form. The result of this for the terminology work in this special sector, besides the recording of word pairs or individual designations, as it is practiced in particular in the technical language field, is the requirement of extracting relatively long textual entries from the original documents so that at a later point in time, for example, important points of an accord can be reproduced in the official and binding version without having to first undertake a time consuming study of the original sources. A lexicographical treatment (condensation of the text, returning parts of speech to the infinitive, etc.) in a manner customary for the production of dictionaries may thus not be undertaken in the existing political and juridical phraseology<sup>1</sup>. In practice, one can only proceed so that the original text entries which are involved are recorded in the corresponding languages while maintaining absolute parallelism between the source and target languages, and a characteristic indexing word (or if required, even several) occurring in the text entry is chosen for the storage of the collected material, which then makes it possible to find the desired text entry within the framework of a card catalog or a similar arranging system.

For reasons of space, in this case the number of indexing words will have to be kept relatively low with a conventional card file, while in the case of a word list or with a printed glossary, one can work with redundancy throughout in order to offer the user a better chance of finding the desired information with a one-time referral where possible.

For a conventional collection of this type, there is naturally the requirement that it must be available in all working language directions if it is to fulfill its purpose. This means a considerable consumption of time for the conversion to the particular source languages (the language spectrum for which terminology is needed directly in the documentation of the language service of the Foreign Office extends at any rate over all official United Nations and Preliminary Law Book languages) as well as for the subsequent alphabetizing.

Obviously, a point has already been reached here at which one must pose the question whether a conventional collection (card catalog, etc.) can still be

---

<sup>1</sup> The following distinction applies within the framework of this article:

Terminology = Individual technical terms, i.e. single and multiple word designations;

Phraseology = Technical language idioms, formulas, including up to entire sentences.



up to the requirements, since its management entails a considerable expense in time and personnel if the number of entries goes into the hundreds of thousands. Above and beyond this, a trend is to be noted at the present time which is leading to an incalculable growth of technical vocabulary in the most diverse fields. The efficient execution of translations, and the work related to them, all the more require an extensive, high quality and reliable vocabulary, which additionally is available in a form which assures the easy, rapid and sure capability of locating words, is suited to directed interrogation in accordance with the most diverse criteria and can be maintained with little expense and without problems at the state of the art.

Today, requirements must be placed on an effective system of stored information retrieval in accordance with set criteria (information retrieval system) for the technical language area (systems for solving documentation problems of the most diverse types are really already successfully in service today in many places), especially as far as rapid and differentiated access are concerned, as well as a fast updating, which can practically be solved only by means of electronic data processing, i.e. by means of a terminological data bank which can offer the corresponding applications range in this regard.

Requirements of this type would be:

- Space saving disposition of the stored information;
- The capability of bringing the material up to date quickly, thereby maximum timeliness at any point in time;
- Fast availability of the stored information, and specifically not only the total information store, but especially partial bodies of it, which satisfy set selection criteria.

Some of the selection capabilities which are in the greatest demand and which the TEAM data bank system mentioned at the outset offers, will be briefly enumerated here:

- According languages and language combinations;
- According to technical fields and technical field combinations;
- According to sources;
- According to parts of speech;
- According to the input date;
- According to quality features;
- According to the person who fed in the information;
- According to the initial letters;
- According to association with a system or piece of equipment;
- According to the existence of definitions;

- According to the existence of synonyms;
- According to abbreviations.

These selections cited here, which represent only a part of those possible, are above and beyond this capable of being made in an interrelated fashion. The search is not really to be made in many cases according to just a single set criterion (perhaps according to a term and its equivalent in the foreign language), but rather a selective inquiry is to be made, i.e. the scope of the answer should be limited at the outset through directed question combinations in order not to burden the person asking the question with material he does not at all need. Selections of this type in almost any combination present no difficulties at all for the TEAM data bank system. It would take too long to go into all of the conceivable combinations here.

However, by way of illustration, here is an example such as could occur in practice:

To be searched out of a four language stock (G/E/F/R) of 100,000 entries are all French entries which were fed in by researcher X prior to January 1st, 1971, and which incorrectly bear the quality marker "ready for print", although they come from the non-obligatory XYZ source.

This task would be accomplished by a medium data processing system of the SIE MENS 4004 family with the corresponding selection program of the TEAM system in less than half an hour and in one run. For example, it could provide a printout through a high speed printer in which the entries were precisely listed and which would meet all requisite criteria.

One should be able to see to an adequate extent in light of the example just given here how flexible and capable a data bank can be in this respect.

Additional requirements, which would not actually come into play for every group of users, for example, would also be:

Clear display of the overall stock or any part of it.

In the case of the Foreign Office it is frequently necessary to make extracts from a vocabular for a certain subject area available to not only its own diplomatic representatives all over the world, but also to the representatives of other governments. The DIGISET photocomposer is ideal for this purpose, which will be discussed in yet more detail at a later point.

Direct access to the total information store via data viewing terminals.

This is certainly at the moment probably the most spectacular form of data bank utilization, and the TEAM system is also completely set up for this. The relatively high financial outlay in the case of small circles of users makes the application of display terminals not yet profitable for the interrogation of terminological data banks in most cases, since the equipment cannot be utilized to full capacity.

The lexicographical utilization, which will be treated in complete detail later, covers so much in the editing field and offers such manifold possibilities for information processing and individual retrieval, that it provides extremely satisfactory results for most requirements, beginning with the rapid compilation of terminology lists via high speed printer for specific subject areas up to completely made-up print manuscripts composed via the DIGISET CRT photocomposing system.

## II. The TEAM Program System

The TEAM program system consists of a series of program elements by which the input, checking, processing and output stages are largely subdivided into independent sections. The data unit is the so-called "entry", i.e. the totality of the information incorporated into a certain concept: designations in the various languages, source citations, definitions, synonyms, etc. Each entry is broken down into 100 so-called "categories", which are individually addressable and (with the exception of categories 00 and 99) can be arbitrarily occupied or released. Each of these categories contains specific information elements which together yield the entry.

The primary recording of the terminology data is accomplished by means of a SIEMENS teletypewriter with an attached perforator. A maximum of 116 characters is available for the input through the use of this teletypewriter, of which 22 alone are diacritical marks, thus a store of characters which permits the recording of the entire Latin alphabet in natural orthography (upper and lower case letters, umlauts, numerals and diacritical marks, as well as the usual punctuation marks, and additionally even numerous special characters). Above and beyond this, the recording of non-Latin scripts (for example, cyrillic) is possible without difficulty, and above all, without any loss of information using a transliteration system recommended by the ISO. The use of other procedures (for example, the application of punched cards, magnetic tape typewriters, or OCR-B typewriters in conjunction with a plain text reader in place of the teletypewriter/perforated tape reader combination, is likewise possible. Recently being used to an increased extent in the TEAM system for recording are the comfortable OCR-B typewriters. The teletypewriters otherwise used for the TEAM data recording also offer the not to be underestimated advantage of an easily readable format, so that the first corrections of typographical errors, etc., can be made early during the recording and prior to the actual computer run, something which means a considerable cost savings.

In the transfer of the perforated tapes to magnetic tape (see Figure 1), the data are checked for formal correctness. Entries detected as faulty are printed out via the high speed printer as a so-called "error report". Since they are not transferred at the same time to magnetic tape, they must be fed in again after being formally corrected.

The formally correct entries are converted to the code of the data processing system and stored on magnetic tape. After they have been numerically sorted, they can either run through a correction program and then be available as cleaned data units for further processing in the TEAM system, or they can be incorporated into an already existing "main data file" by means of a program for updating them.

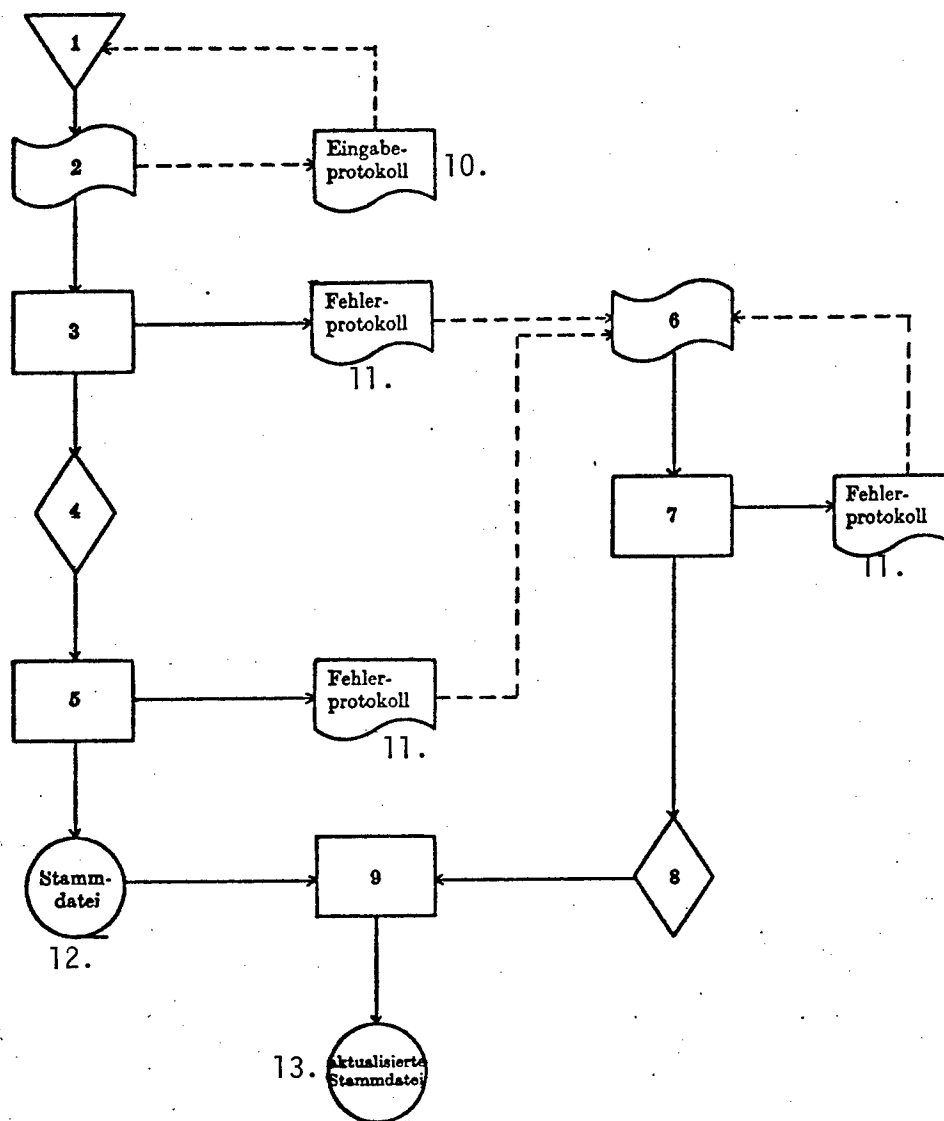


Figure 1. Data flow chart: Generating and updating a main data file.

Key: 1. Manuscript; 2. Perforated tapes (basic entries and corrections); 3. Perforated tape transfer to magnetic tape; 4. Numerical sorting; 5. Correction program; 6. Perforated tapes (correction entries); 7. Perforated tape transfer to magnetic tape; 8. Numerical sorting; 9. Updating program; 10. Input record; 11. Error report record; 12. Main data file; 13. Updated main data file.

As soon as the groundwork for the "main data file" has been laid, all subsequent operations with this basic stock can be broken down into two work areas:

- A) Updating and expanding the basic stock;
- B) Providing the user with the stock or a part of the stock (retrieval).

Besides the already mentioned program elements for producing the main data file, there are a number of other programs available for the working area which can be used as required. Included here, among others, is a program which detects double or multiple entries, and based on set guidelines, collates them and reports them as "doublet suspected", as well as a program with which corrections can be made for individual categories in an unlimited sequence of entries via punched cards.

The other work area is exclusively oriented towards the needs and desires of the data bank user. The TEAM program system attempts to do justice to the quite diverse types of requirements as regards the terminology to be fed out, as well as the required supplemental information (sources, definitions, etc.), through several selection programs. The achievements of this program have already been reported above.

One of these selection programs fulfills yet another important function by creating the prerequisites for the alphabetical sorting capability in the individual languages. It produces so-called "sorting concepts"<sup>1</sup> for those languages in which the term designations are to be alphabetically sorted. Based on these sorting concepts, the terms can be sorted over any number of places in accordance with the alphabetizing rules of the particular language. By this means, on one hand the correct alphabetical sequence is achieved for the Latin, as well as for the non-Latin alphabets, and on the other hand, special characters such as ß, ä, ö and ü in German, ñ and ll in Spanish, etc., can be placed in their own order in the language concerned. Such subtleties, which are indispensable, primarily in lexicographical work, can in no case be realized with the standard sorting and mixing programs usually maintained in computer centers without preliminary sorting concept generation.

The appropriate TEAM program permits the formation of several sorting concept types, of which the "through sorting", "sorting according to key words" and "sorting according to nests" are to be briefly discussed.

The necessity of these program variants becomes clear if one keeps in mind that with present day capacities for computer information processing, for example in dictionary production, computer controlled photocomposition takes the place of conventional composing, where the photocomposing system provides finished manuscripts after the corresponding programs have completed the requisite preliminary work.

#### Through Sorting

In this procedure, all punctuation marks and special characters, as well as gaps between words, remain out of consideration.

---

<sup>1</sup> Sorting concept = Criterion for the sorting sequence.

Example:

Eis  
Eisalaun  
Eisen  
Eisenabfall  
Eisen, altes  
Eisenbahn  
Eisenbahnabteil  
Eisenbahn, elektrische  
Eisenbahnnetz  
Eisen, kristallinisches  
Eisenlegierung  
Eisen, meteorisches  
Eisentfernung  
Eisenträger  
Eisessig

The above sorting sequence is standardized for the German area as DIN 5007 [German Industrial Standard] for alphabetizing names.

#### Sorting According to Key Words

In this method, gaps or spaces between words appearing after a word are taken into account, so that paragraph formation (i.e. the incorporation of several entries with the same key word into one paragraph) is made possible in the print format.

Example:

Key Word Sorting	Paragraph Formation
Eis	Eis
Eisalaun	Eisalaun
Eisen	Eisen; altes ~;
Eisen, altes	kristallinisches ~;
Eisen, kristallinisches	meteorisches ~
Eisen, meteorisches	Eisenabfall
Eisenabfall	Eisenbahn; elektrische ~
Eisenbahn	Eisenbahnabteil
Eisenbahn, elektrische	Eisenbahnnetz
Eisenbahnabteil	Eisenlegierung
Eisenbahnnetz	Eisentfernung
Eisenlegierung	Eisenträger
Eisentfernung	Eisessig
Eisenträger	
Eisessig	

#### Sorting According to Nests

The word elements are arranged in such a fashion here that they can be incorporated into nests in a subsequent processing program for the photocomposition. A "nest" is understood to be a collection of all idioms and compounds associated with a basic word in a closed, internally ordered block. In the case of nest sorting, affiliation with a nest has predominance over the alphabetical ordering of the overall stock (see below: Eisenträger appears before Eisenbahn!). The nest configuration offers even more of a savings in space than paragraph formation, but at the same time is less clearly arranged.

Example:

#### Nest Sorting

Eis  
Eisalaun  
Eisentfernung  
Eisessig  
Eisen  
Eisen, altes  
Eisen, kristallinisches  
Eisen, meteorisches  
Eisenabfall  
Eisenlegierung  
Eisenträger  
Eisenbahn  
Eisenbahn, elektrische  
Eisenbahnabteil  
Eisenbahnnetz

#### Nest Formation

Eis; ~alaun; ~entfer-  
nung; ~essig  
Eisen; altes ~; kristalli-  
nisches ~; meteori-  
sches ~; ~abfall; ~le-  
gierung; ~träger  
Eisenbahn; elektrische ~;  
~abteil; ~netz

The selection of the appropriate sorting procedure is made taking into account the nature and quantity of the word material concerned, as well as the goals of the circle of users. The thoroughly compact arrangement in the case of the sorting last cited above with subsequent nest formation presents an opportunity, for example, for more extensive bilingual dictionaries, in which the number of pages should remain within a reasonable framework.

#### From the Main Data File to the Finished Printing Manuscript

The TEAM program system is also quite flexible in the output branch. The high speed printer is particularly well suited to putting out editing and discussion manuscripts, as well as short term word lists (for example, for a special translation task). There are several programs in existence at the present time within the framework of the TEAM system, which permit output through library high speed printers developed especially for documentation purposes by Siemens AG, with upper and lower case letters and diacritical marks, as well as via conventional high speed printers (with the corresponding conversion of the information). In these programs, a choice can be made between any of one- to five-column page proofs, as well as horizontal and vertical configurations. Through appropriate parameter input, the output sequence which should be employed can be determined. Pagination and current column headings are produced and inserted at set points by the blank line program. Above and beyond this, there is the capability of selecting between on-line printout or producing a magnetic tape for off-line printing. Furthermore, it can be determined which of the parts prepared for printing should actually be listed, and indeed, according to the page number (for example, from page 47 to page 123), or, in the case of an alphabetically sorted stock, according to the initial letters of the source language (for example, from A to LZ).

The photo-composing system actually represents a far more convenient output medium. Several programs were written for this purpose which prepare the entries stored on magnetic tape for photo-composition and transfer them to input tapes for the photo-composing system.

## transit

des voies de transit / индивидуальные пошлыны и сборы за использование транзитных путей; misuse of the - routes IV 11 / Mißbrauch der Transitwege / abus des voies de transit / злоупотребление транзитных путей

travel, direct - to and from the Western Sectors of Berlin IV 11 / direkte Durchreise von und nach den Westsektoren Berlins / le voyage direct à destination et en provenance des secteurs occidentaux de Berlin / прямой проезд в западные сектора Берлина и из них; vouchers for - and tours IV 11 / Gutscheine für Reisen und Rundreisen / bons pour des voyages / ваучеры для поездок и туров

The TEAM-DIGIS program for output through the DIGISET CRT photo-composing system.

Terminology and phraseology of the Four Power Accord on Berlin.

## V

vehicle, individual -s (persons) IV 11 / individuelle Transportmittel (Personen) / des véhicules individuels (personnes) / индивидуальные средства транспорта  
vouchers for travel and tours IV 11 / Gutscheine für Reisen und Rundreisen / bons pour des voyages / ваучеры для поездок и туров

Two-column with pagination, running column headings and locating aids, medium-face and fine Digi-Antiqua type face, Rossiya fine, seven-point with one-point space. Horizontal arrangement of the languages (English, German, French, Russian) with paragraph formation.

### voie de transit IV 11

D: Transitweg *m*

E: transit route

R: транзитный путь

### voie, abus des -s de transit IV 11

D: Mißbrauch der Transitwege

E: misuse of the transit routes

R: злоупотребление транзитных путей

### voie, l'entretien de -s ... adaptées à cette circulation IV 11

D: Instandhaltung entsprechender Wege

E: maintenance of adequate routes

R: поддержание соответствующих путей

Single column with pagination, medium face and fine Digi-Antiqua type and Rossiya fine, eight-point with one-point space, vertical arrangement of the languages (French, German, English, Russian).

### документ, проверка пломб и сопроводительных -ов IV 11

D: Prüfung der Plomben und Begleitdokumente

E: inspection of seals and accompanying documents

F: contrôle des plombs et des documents d'accompagnement

### должностной, конституционный или - акт IV 11

D: Verfassungs- oder Amtakt

E: constitutional or official act

F: acte constitutionnel ou officiel

### западный сектор IV 11

D: Westsektor *m*

E: Western Sector

F: secteur occidental

### злоупотребление *n* IV 11

D: Mißbrauch *m*

E: misuse *n*

F: abus *m*

### злоупотребление транзитных путей IV 11

D: Mißbrauch der Transitwege

E: misuse of the transit routes

F: abus des voies de transit

Two-column with pagination, running column headings and locating aids, fine Digi-Antiqua and Rossiya fine type, seven-point with one-point spaces. Vertical arrangement of the languages (Russian, German, English, French).

Figure 2 Examples of formatting possibilities with photo-composition.

Besides the pure textual data which should appear in the manuscript, these contain all of the control instructions necessary for the composition, so that a completely made-up one or multi-column manuscript appears on film or



paper, in which no more changes have to be made at all. The manifold possibilities of the typographical formatting are clearly seen from the illustrations (see Figure 2).

Output via photo-composing systems is today already so financially advantageous that it is completely competitive on a cost basis with on-line output via high speed printers. Based on the substantially better readability and handiness, as well as the considerably reduced volume of paper, photocomposition is far superior to the high speed printer though. In the case of non-Latin alphabets, the user cannot generally be expected to read a high-speed printer list with transliterated or completely transcribed text, so that here only output via computer composition from the outset appears acceptable.

Two additional utilization possibilities of a data bank should be briefly noted in passing at this point, which the TEAM system likewise provides, specifically, textually referenced interrogation of the terminology data bank in batch operation (by means of which a translator can query the data stock for terms in the sequence in which they occur in the translation being worked on, or of course, also alphabetically) and computer assisted language training, in which a part of the teaching material is computer produced through the corresponding programs, in which case the computer determination of the particular minimum vocabulary for a technical field, as well as the computer generation of translation aids for technical texts predominate.

### III. A Practical Demonstration of Structuring a Terminological and Phraseological Data Bank Using the Lexical Material from the Four Power Accord of September 3rd, 1971.

The task consisted in recording the terminology and phraseology in the German, English, French and Russian languages contained in a part of the Four Power Accord and marked by coworkers of the language service of the Foreign Office, as well as in reproducing it in a meaningful and clearly understandable form, in which case each of the four languages should be the source language on one occasion. The source material consisted of a series of index cards giving a language translation from German to English, and also consisted of the French and Russian texts of the accord which were provided with underlinings to mark the corresponding points in the text.

#### Problem Analysis

The first thing to be investigated using the source material was in which typical configurations deviating from the standard form did the information to be recorded appear, and to what extent this could be brought into agreement with the existing input conventions of the TEAM system. It proved to be the case that the TEAM system, which is basically designed for terminological data banks with simple word equivalents, was also completely suited to phraseology of a political and juridical character structured in a quite different manner.

The effort was made with longer text entries to bring in as many meaningful index words as possible. In this way, the result was actually a certain redundancy in the end product, however, on the other hand also corresponding convenience for the user.

Example:            Prüfung der Plomben und Begleitdokumente  
                      appears under    Prüfung  
    Plombe  
    Begleitdokument  
                      and the English equivalent  
                      Inspection of seals and accompanying documents  
                      appears under    Inspection  
    Seal  
    Document.

Only for a special case was it necessary to have an inherent rule, specifically for longer phrases with several index words in which it was not possible to reduce the indexing words to a basic form (= key word form) by simply removing the inflected ending. The solution to this problem is explained in more detail in the following section.

#### Establishing the Input Format

The input format is the form and sequence in which the individual information units of an entry are stored. In this case, the entry is broken down into small units with rigidly defined contents, by which means the effect is achieved of making all individual information units belonging to a total information block accessible. This is especially important for corrections and selection.

The individual information categories which were used for the trial were defined as follows:

Category	Contents	Explanation
03	d	= Ready for printing (further differentiation can be selected according to need, for example, v = obligatory; A = working concept, etc.)
04	0972	= Recording date (month and year)
05	B124	Code number of the terminology worker responsible or the person who fed the data in.
06	V1971	= 1971 accord (in this case, an arbitrarily selected code for the Four Power Accord).

The information elements enumerated to this point remain constant for the entire contract and need only to be recorded once as a so-called "lead-in" at the start of the actual language categories. In the subsequent processing, the perforated tape transfer program handles the assigning of these constant information units to each individual entry.

Additional categories for the standard case, i.e. for simple word equivalents and multiple word designations, in which there were no lemmatization problems:

Category	Explanation
00	Address of the entry (alphanumeric)
10	German designation
11	Part of speech of the German designation (only for the case of one word designations)
20	English designation
21	Part of speech of the English designation
30	French designation
31	Part of speech of the French designation
50	Russian designation
51	Part of speech of the Russian designation.

With multiple word designations and phrases, control characters were specified for the generation of inversions (\*), for the separation of inflected endings ('), as well as for prohibition of inversion (..).

00	XY3033	
10	Gebühren	
11	f.,pl.	Figure 3. Perforation example
20	tolls	
21	n.,pl.	The character 99< means that the information unit with the XY3033 address is terminated.
30	péages	
31	m.,pl.	The Russian text is transliterated and with DIGISET output, is again back-transliterated.
50	pošlina	
51	f.	
	99</	

It was possible through this type of recording to have the uninflected index word appear as the key word in the later printout, followed by the entire phrase in its natural word sequence and the corresponding foreign language versions.

#### Data Preparation and Recording

The preparation of the word material was carried out by the language service of the Foreign Office, and in this case, in two different forms:

1. The German/English equivalents, as already mentioned, were already down in card index form. For this reason, an obvious step was to draw directly on this card index for the data recording.

Preparation of this type in card index form, or completely manual alphabetizing is naturally unnecessary; in the case of multilingual data (as in the present case with an accord existing in four language versions). however, it is recommended that an easily traceable relationship be established between text locations in the individual languages, perhaps in the form of numeration as was the case with the remainder.

2. In the Russian and French original text of the accord, the corresponding points were designated by underlining and assigned by numeration to the German and English text locations established in the index card file.

The recording of the phraseology prepared in this fashion was accomplished via the Siemens teletypewriter on six-channel perforated tape, which was mentioned earlier.

Figure 4. Perforation example for it:

```
00 XY3032
10 ..andere kleine *Gebiete
20 ..other small *areas
30 ..autres *parcelles
99</
```

An input record (Figures 3, 4 and 5 are parts of it) was produced during the recording, and was subsequently proofread. In most cases, one-time proofreading immediately after recording is adequate, and makes it possible to detect incorrect points prior to the actual processing through the computer, and where the perforated tapes are transferred to magnetic tape, simultaneously

Figure 5. A complete entry looks as follows when punched in:

```
00 XY3084
10 Vereinbarungen und Beschlüsse der vier Mächte aus der
   Kriegs- und Nachkriegszeit
17 Vereinbarungen;f.,pl.: Beschluß;m.; Macht;f.;
   Kriegszeit;f.; Nachkriegszeit;f.
20 wartime and *post-war *arrangements and *decisions
   of the Four *Powers
30 ..les *accords et *décisions des quatre *puissances
   au temps de la *guerre et de l'*après-guerre
50 soglašenija i rešenija četyreh deržav voennogo i
   poslevoennogo vremeni
57 soglašenje;n.; rešenje;n.; deržava;f.; voennyj;adj.;
   poslevoennyj;adj.; vremja;n.
99</
```

Figure 6. DIGISET sample for the preceding example:

Key word, "Beschluss"

**Beschluß m** [V1971] Vereinbarungen und Beschlüsse der vier Mächte aus der Kriegs- und Nachkriegszeit  
*E:* wartime and post-war arrangements and decisions of the Four Powers  
*F:* les accords et décisions des quatre-puissances au temps de la guerre et de l'après-guerre  
*R:* соглашения и решения четырех держав военного и послевоенного времени

feeding in the requisite correction entries. Where necessary, corrections can of course be inserted in any later phase in any quantity.

### Special Cases

In the case of multiple word designations and phrases, in which an index word appears in an inflected form which cannot be traced back to the basic form by splitting off the inflected ending, an additional information category was adopted as an aid which contains the index words concerned in the basic form (category -7). This form occurs not all too frequently in German, English and French, however, quite often in Russian.

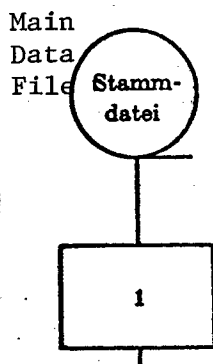
### Machine Processing

The machine processing was carried out with the programs of the TEAM system using a computer facility of the SIEMENS 4004 family.

The input data consisted of four-language entries on perforated tape, along with the associated corrections, in a mixed sequence.

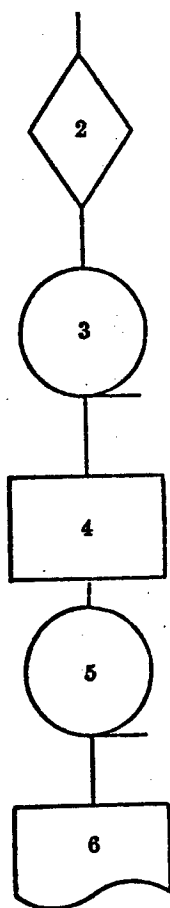
Until the production of the main data file, the course taken was that already described earlier and provided by the TEAM system for the structuring of a data bank. The contents of the main data file were printed out via a high speed printer with an available printout program to check and sort the recorded stock.

The subsequent data flow chart then looked as follows:



1. With respect to the desired alphabetizing, the sorting concept formation was first undertaken for the individual languages, as well as the generation of inversion and synonym entries.

The sorting was accomplished here as so-called key word sorting.



2. The total stock which was provided with sorting concepts was then immediately sorted alphabetically with a service program.
3. The resulting magnetic tape, which contains the data in alphabetical order, serves as the basis for the output branch of the TEAM program system, in which an output is desired in alphabetical form.

For the sake of completeness, it should be noted that a so-called doublet run can be undertaken in this phase, in which double and multiple entries are excluded and so-called "doublet suspected" entries are reported.

4. The subsequent Digiset preparation program processes the alphabetically sorted stock through the corresponding parameter input for the photo-composition.
5. The resulting magnetic tape controls the Digiset photo-composing system directly, and, besides the format parameters (see Figure 8), contains all of the control characters (for the script, type size, medium and light face type, etc.) necessary for composition.
6. The letters are formed electronically on a cathode ray tube in the photo-composing system and transferred to film (or also to paper) by means of an optical system.

The resulting product is a completely made-up printing manuscript.

The course depicted here had to be carried out once for each source language with the specific, different parameters. Furthermore, the exterior formatting was varied slightly each time in order to show the flexibility of the system (Figure 2).

#### Data Output

All of the already described capabilities of the TEAM program system were available for the data output. Chosen for the word material of the Foreign Office which is treated here was the most demanding version (DIGISET photo-composition, Figure 2), with a multiplicity of individual formatting possibilities. The type style can be configured in an extremely flexible fashion. Through the appropriate parameter input (cf. Figure 8), any desired format is to automatically achieved; in this case, at the time of the recording, no decision is to be made as to what order the data should appear in during the subsequent printing.

DIGIS/KAT	200609100930095099
DIGIS/SPRA	H
DIGIS/ABS	JA
DIGIS/KYR	JA
DIGIS/TIT	JA
DIGIS/IWE	JA,12
DIGIS/GROE	7
DIGIS/DUR	6
DIGIS/ZEIL	0120
DIGIS/ZLBR	1060
DIGIS/SPBR	1177

Figure 8. Parameters in the case of the Digiset preparation program.

RUNDREISE F.	XY3057
E: TOUR N.	D 0972
F: VOYAGE M.	B124
R: TUR M.	
V1971	
GUTSCHEINE FUER *REISEN UND *RUNDREISEN	XY3058
E: VOUCHERS FOR *TRAVEL AND *TOURS	D 0972
F: BONS POUR DES *VOYAGES	B124
R: VAUCHERY DLJA POEZDOK I *TUROV	
V1971	
..ANGELEGENHEITEN DER *SICHERHEIT UND DES	XY3059
*STATUS	D 0972
E: ..MATTERS OF *SECURITY AND *STATUS	B124
F: ..QUESTIONS DE *SECURITE ET DE	
*STATUT	
R: VOPROSY BEZOPASNOSTI I *STATUSA	
V1971	

Figure 9. Excerpt from the high-speed printer list of the entire stock (numerical).

Here, the following arranging and composing instructions are established: category sequence (English/German/French/Russian), language configuration (horizontal), paragraph formation, cyrillic type font, column heading formation, exchange of initials, type size (here, seven point), space (here one point), number of lines, line width, column width.

Printout via high speed printers is particularly suited to less demanding goals (for example, for a technical glossary as a short term working aid for internal use). Two versions are available for this: the normal high speed printer, which supplies only upper case type and a few special characters, and the library high-speed printer, which has an increased stock of characters, and upper and lower case type, and also provides for the reproduction of diacritical marks.

The high-speed printer was used in the case at hand here for the numerical listing of the recorded stock (see Figure 9).

It is again to be especially noted here that any information can be suppressed as required in all output programs (for example, a two language list can be fed out from a four language stock; or only the purely linguistic information, without source citations, subject areas, miscellaneous code numbers, etc., can be printed out). In this way, the result for the user is the capability of dispensing with any ballast at all in information retrieval and extracting only the salient part, thus the information which he actually needs, as required, from the stock stored in the data bank.

A further point which perhaps needs mentioning is the scope of the vocabulary. In this trial, a stock of around 100 basic entries was processed, which is quite small. The managing of more extensive stocks entails no additional problems in data recording, other than the time factor. This is really the specific advantage of a data bank, i.e. being able to operate almost as quickly with a multiple of this stock. To illustrate this, it should be said that the alphabetical sorting of 60,000 entries with four working tapes and a memory occupancy of 100 KB, including the clearing of the tapes, the reading in of the control cards and the loading of the program (so-called "set-up" time) last only 20 minutes.

In conclusion, one can probably say without reservation that the TEAM data bank system has proven its performance capability and flexibility convincingly in the field of phraseology. Since it appears to be reasonable to also make the overall final product accessible to the readers of LEBENDE SPRACHEN, the resulting glossary of the "Terminology and Phraseology of the Four Power Accord on Berlin" translated from German into English, French and Russian appears at the end of this article. It was produced with a DIGISET photo-composing system with fully automatic generation of the page proof needed for the page format of LEBENDE SPRACHEN.

R. Schmidt/O. Vollnhals  
Siemens AG, Munich



# THE TERMINOLOGY AND PHRASEOLOGY OF THE FOUR POWER ACCORD ON BERLIN

<b>Terminologie und Phraseologie des Viernächte-Abkommens über Berlin</b>	Республикой Германии будут поддерживаться	pas entravée
<b>Abfahrt, vor der - verplombte Transportmittel [V1971]</b>	Abschaffung von Vereinbarungen und Abmachungen [V1971]	R: транзитное сообщение - будет происходить без помех
E: conveyances sealed before departure	E: extension of agreements and arrangements	Behörde, nachfolgende Vereinbarungen und Regelungen zwischen den zuständigen deutschen -n [V1971]
F: des véhicules plombés avant le départ	R: расширение соглашений и договоренностей	E: the consequent agreements and arrangements between the competent German authorities
R: опломбированные перед отправлением перевозочные средства	Außenhandelsvereinigung f [V1971]	F: les accords et arrangements entre les autorités allemandes competentes qui s'ensuivent
<b>Abgabe, individuelle Gebühren und -n für die Benutzung der Transitwege [V1971]</b>	F: foreign trade association	R: последующие соглашения и урегулирования компетентных немецких властей
E: individual tolls and fees for the use of the transit routes	R: внешнеторговое объединение	Behörde, ständige Verbindungs- [V1971]
F: péages et taxes individuels pour l'usage des voies de transit	Außenhandelsvereinigung, Büro der sowjetischen -en in den Westsektoren Berlins [V1971]	E: permanent liaison agency
R: индивидуальные пошлины и сборы за использование транзитных путей	E: Office of Soviet Foreign Trade Associations in the Western Sectors of Berlin	F: office de liaison permanent
<b>Abgabe, Kompensation für -n, Gebühren und andere Kosten [V1971]</b>	F: Bureau des Associations Soviétiques de Commerce Extérieur dans les secteurs occidentaux de Berlin	R: постоянный орган по связи
E: compensation for fees and tolls and for other costs	R: Бюро Советских Внешнеторговых Объединений в западных секторах Берлина	Berlin, das Verhältnis zwischen den Westsektoren -s und der BRD [V1971]
F: compensation pour les péages et taxes et pour les autres frais	Ausübung, Akt in - unmittelbar	E: the relationship between the Western Sectors of Berlin and the FRG
R: компенсация за сборы, пошлины и другие расходы	E: act in exercise of direct state authority	F: la relation entre les secteurs occidentaux de Berlin et la RFA
<b>Abkommen, Viernächte- [V1971]</b>	F: acte accompli dans l'exercice d'une compétence étatique directe	R: отношение между западными секторами Берлина и Федеративной Республикой Германии
E: quadripartite agreement	R: акт в осуществление непосредственной государственной власти	Berlin, die Bindungen zwischen den Westsektoren -s und der BRD werden aufrechterhalten [V1971]
F: accord quadripartite		E: the ties between the Western Sectors of Berlin and the FRG will be maintained
R: четырехстороннее соглашение		
<b>Abmachung, Ausdehnung von Vereinbarungen und -en [V1971]</b>		

# THE TERMINOLOGY AND PHRASEOLOGY OF THE FOUR POWER ACCORD ON BERLIN

<b>E:</b> extension of agreements and arrangements <b>F:</b> extension d'accords et arrangements <b>R:</b> распространение соглашений и договоренностей <b>Akt in Ausübung unmittelbarer Staatsgewalt [V1971]</b> <b>E:</b> act in exercise of direct state authority <b>F:</b> acte accompli dans l'exercice d'une compétence étatique directe <b>R:</b> акт в осуществление непосредственной государственной власти <b>Amtsakt, Verfassungs- oder [V1971]</b> <b>E:</b> constitutional or official act <b>F:</b> acte constitutionnel ou officiel <b>R:</b> конституционный или должностной акт <b>aufrechterhalten, die Bindungen zwischen den Westsektoren Berlins und der BRD werden [V1971]</b> <b>E:</b> the ties between the Western Sectors of Berlin and the FRG will be maintained <b>F:</b> les liens entre les secteurs occidentaux de Berlin et la RFA seront maintenus <b>R:</b> связи между западными секторами Берлина и Федеративной	<b>Autobus, durchgehende Züge und -se [V1971]</b> <b>E:</b> through trains and buses <b>F:</b> des trains et autocars directs <b>R:</b> безостановочные поезда и автобусы <b>Bedienstete p/ [V1971]</b> <b>E:</b> personnel <i>n</i> <b>F:</b> personnel <i>m</i> <b>R:</b> сотрудники <i>m pl</i> <b>Bedienstete, Zuweisung von -n an das Generalkonsulat [V1971]</b> <b>E:</b> assignment of personnel to the Consulate General <b>F:</b> affectation de personnel au Consulat général <b>R:</b> направление сотрудников на работу в Генеральное консульство <b>Begleitdokument, Prüfung der Plomben und -e [V1971]</b> <b>E:</b> inspection of seals and accompanying documents <b>F:</b> contrôle des plombs et des documents d'accompagnement <b>R:</b> проверка пломб и сопроводительных документов <b>Behinderung, der Transitverkehr wird - ohne -en sein [V1971]</b> <b>E:</b> transit traffic will be ... unimpeded <b>F:</b> la circulation en transit - ne sera	<b>F:</b> les liens entre les secteurs occidentaux de Berlin et la RFA seront maintenus <b>R:</b> связи между западными секторами Берлина и Федеративной Республики Германии будут поддерживаться <b>Berlin, die Kommunikationen zwischen den Westsektoren -s und Gebieten, die -n angrenzen [V1971]</b> <b>E:</b> the communications between the Western Sectors of Berlin and areas bordering on - <b>F:</b> les communications entre les secteurs occidentaux de Berlin et les régions limitrophes de ces secteurs <b>R:</b> сообщения между западными секторами Берлина и пограничными с этими секторами районами <b>Berlin, die Verbindungen der Westsektoren -s nach außen [V1971]</b> <b>E:</b> the external communications of the Western Sectors of Berlin <b>F:</b> les communications des secteurs occidentaux de Berlin avec l'extérieur <b>R:</b> внешние связи западных секторов Берлина <b>Berlin, direkte Durchreise von und nach den Westsektoren -s [V1971]</b> <b>E:</b> direct travel to and from the
<b>Western Sectors of Berlin</b> <b>F:</b> le voyage direct à destination et en provenance des secteurs occidentaux de Berlin <b>R:</b> прямой проезд в западные сектора Берлина и из них <b>Beschluß m [V1971]</b> Vereinbarungen und Beschlüsse der vier Mächte aus der Kriegs- und Nachkriegszeit <b>E:</b> wartime and post-war arrangements and decisions of the Four Powers <b>F:</b> les accords et décisions des quatre puissances au temps de la guerre et de l'après-guerre <b>R:</b> соглашения и решения четырех держав военного и послевоенного времени <b>Bestandteil m (konstitutiver Teil) [V1971]</b>	<b>Berlin</b> <b>DDR, das Territorium der [V1971]</b> <b>E:</b> the territory of the GDR <b>F:</b> le territoire de la RDA <b>R:</b> территория Германской Демократической Республики <b>Dienste, Luftfracht- [V1971]</b> <b>E:</b> air freight services <b>F:</b> service de fret <b>R:</b> услуги по воздушным перевозкам <b>direkte Durchreise von und nach den Westsektoren Berlins [V1971]</b> <b>E:</b> direct travel to and from the Western Sectors of Berlin <b>F:</b> le voyage direct à destination et en provenance des secteurs occidentaux de Berlin <b>R:</b> прямой проезд в западные	<b>R:</b> обмен территория <b>Gebühren p/ [V1971]</b> <b>E:</b> tolls <i>pl</i> <b>F:</b> péages <i>m pl</i> <b>R:</b> пошлина <i>f</i> <b>Gebühren, individuelle - und Abgaben für die Benutzung der Transitwege [V1971]</b> <b>E:</b> individual tolls and fees for the use of the transit routes <b>F:</b> péages et taxes individuels pour l'usage des voies de transit <b>R:</b> индивидуальные пошлины и сборы за использование транзитных путей <b>Gebühren, Kompensation für Abgaben, - und andere Kosten [V1971]</b> <b>E:</b> compensation for fees and tolls and

# THE TERMINOLOGY AND PHRASEOLOGY OF THE FOUR POWER ACCORD ON BERLIN

<i>E</i> : constituent part <i>F</i> : élément constitutif <i>R</i> : составная часть	сектора Берлина и из них	for other costs <i>F</i> : compensation pour les péages et taxes et pour les autres frais <i>R</i> : компенсация за сборы, пошлины и другие расходы
<i>Bestimmung</i> f [V1971] die Bestimmungen des Grundgesetzes ... sind suspendiert worden <i>E</i> : the provisions of the Basic Law ... have been suspended <i>F</i> : les dispositions de la loi fondamentale ... ont été suspendues <i>R</i> : положения основного закона ... были приостановлены в своем действии	Durchgangsverkehr m [V1971] <i>E</i> : through traffic <i>F</i> : circulation directe <i>R</i> : сквозной проезд	Generalconsulat, Zuweisung von Bediensteten an das [V1971] <i>E</i> : assignment of personnel to the Consulate General <i>F</i> : affectation de personnel au Consulat général <i>R</i> : направление сотрудников на работу в Генеральное консульство
betreffend, das -e Gebiet (Berlin) [V1971] <i>E</i> : the relevant area <i>F</i> : la région correspondante <i>R</i> : соответствующий район	Durchgehende Züge und Autobusse [V1971] <i>E</i> : through trains and buses <i>F</i> : des trains et autocars directs <i>R</i> : безостановочные поезда и автобусы	Grundgesetz n [V1971] die Bestimmungen des Grundgesetzes ... sind suspendiert worden <i>E</i> : the provisions of the Basic Law ... have been suspended <i>F</i> : les dispositions de la loi fondamentale ... ont été suspendues <i>R</i> : положения основного закона ... были приостановлены в своем действии
Bindung, die -en zwischen den Westsektoren Berlins und der BRD werden aufrechterhalten [V1971] <i>E</i> : the ties between the Western Sectors of Berlin and the FRG will be maintained <i>F</i> : les liens entre les secteurs occidentaux de Berlin et la RFA seront maintenus <i>R</i> : связи между западными секторами Берлина и Федеративной Республикой Германии будут поддерживаться	Durchreise, direkte - von und nach den Westsektoren Berlins [V1971] <i>E</i> : direct travel to and from the Western Sectors of Berlin <i>F</i> : le voyage direct à destination et en provenance des secteurs occidentaux de Berlin <i>R</i> : прямой проезд в западные сектора Берлина и из них	Gutscheine für Reisen und Rundreisen [V1971] <i>E</i> : vouchers for travel and tours <i>F</i> : bons pour des voyages <i>R</i> : ваучеры для поездок и туров
BRD, das Verhältnis zwischen den Westsektoren Berlins und der [V1971] <i>E</i> : the relationship between the Western Sectors of Berlin and the FRG <i>F</i> : la relation entre les secteurs occidentaux de Berlin et la RFA <i>R</i> : отношение между западными секторами Берлина и Федеративной Республикой Германии	einreisen, in diese Gebiete -de Personen [V1971] <i>E</i> : persons entering these areas <i>F</i> : personnes se rendant dans ces régions <i>R</i> : лица, въезжающие в эти районы	hinreichende Verdachtsgründe dafür vorliegen, daß ... <i>E</i> : cases in which there is sufficient reason to suspect that ... <i>F</i> : des cas ... où il y aura raison suffisante de soupçonner que ... <i>R</i> : когда имеется достаточное основание подозревать, что ...
BRD, die Bindungen zwischen den Westsektoren Berlins und der - werden aufrechterhalten [V1971] <i>E</i> : the ties between the Western Sectors of Berlin and the FRG will be maintained <i>F</i> : les liens entre les secteurs occidentaux de Berlin et la RFA seront maintenus <i>R</i> : связи между западными секторами Берлина и Федеративной Республикой Германии будут поддерживаться	entsprechend, Instandhaltung -er Wege [V1971] <i>E</i> : maintenance of adequate routes <i>F</i> : l'entretien de voies ... adaptées à cette circulation <i>R</i> : поддержание соответствующих путей	Identifizieren v [V1971] Personen, die als Transitreisende identifiziert sind <i>E</i> : persons identified as through travellers <i>F</i> : les personnes identifiées comme voyageurs directs <i>R</i> : лица, идентифицированные как транзитные пассажиры
	Gebiet, das betreffende (Berlin) [V1971] <i>E</i> : the relevant area <i>F</i> : la région correspondante <i>R</i> : соответствующий район	Instandhaltung entsprechender Wege [V1971] <i>E</i> : maintenance of adequate routes

# THE TERMINOLOGY AND PHRASEOLOGY OF THE FOUR POWER ACCORD ON BERLIN

<b>BRD, sonstige staatliche Organe der IV1971</b>	<b>E:</b> the communications between the Western Sectors of Berlin and areas bordering on --	<b>F:</b> l'entretien de voies -- adaptées a cette circulation
<b>E:</b> other state bodies of the FRG	<b>F:</b> les communications entre les secteurs occidentaux de Berlin et les régions limitrophes de ces secteurs	<b>R:</b> поддержание соответствующих путей
<b>F:</b> d'autres organes étatiques de la RFA	<b>R:</b> сообщения между западными секторами Берлина и пограничными с этими секторами районами	<b>Klarstellungen und Interpretationen IV1971</b>
<b>R:</b> другие государственные учреждения Федеральной Республики Германии	<b>Gebiet, in diese -- einreisende Personen IV1971</b>	<b>E:</b> clarifications and interpretations
<b>Büro der sowjetischen Außenhandelsvereinigungen in den Westsektoren Berlins IV1971</b>	<b>E:</b> persons entering these areas	<b>F:</b> éclaircissements et interprétations
<b>E:</b> Office of Soviet Foreign Trade Associations in the Western Sectors of Berlin	<b>F:</b> personnes se rendant dans ces régions	<b>R:</b> разъяснения и толкования
<b>F:</b> Bureau des Associations Soviétiques de Commerce Extérieur dans les secteurs occidentaux de Berlin	<b>R:</b> лица, въезжающие в эти районы	<b>klein, andere -- Gebiete IV1971</b>
<b>R:</b> Бюро Советских Внешнеторговых Объединений в западных секторах	<b>Gebietsaustausch m IV1971</b>	<b>E:</b> other small areas
<b>bordering on --</b>	<b>E:</b> exchange of territory	<b>F:</b> autres parcelles
<b>F:</b> les communications entre les secteurs occidentaux de Berlin et les régions limitrophes de ces secteurs	<b>F:</b> échange de territoire	<b>R:</b> другие небольшие районы
<b>R:</b> сообщения между западными секторами Берлина и пограничными с этими секторами районами	<b>F:</b> les accords et arrangements entre les autorités allemandes compétentes qui s'ensuivent	<b>Kommunikation, die --en zwischen den Westsektoren Berlins und Gebieten, die -- angrenzen IV1971</b>
<b>Kompensation für Abgaben, Gebühren und andere Kosten IV1971</b>	<b>R:</b> последующие соглашения и урегулирования компетентных немецких властей	<b>E:</b> the communications between the Western Sectors of Berlin and areas
<b>E:</b> compensation for fees and tolls and for other costs	<b>Nachkriegszeit f IV1971</b>	<b>F:</b> legal view
<b>F:</b> compensation pour les péages et taxes et pour les autres frais	<b>und Beschlüsse der vier Mächte aus der Kriegs- und Nachkriegszeit</b>	<b>R:</b> правовой взгляд
<b>R:</b> компенсация за сборы, пошлины и другие расходы	<b>E:</b> wartime and post-war arrangements and decisions of the Four Powers	<b>Rechtsposition f IV1971</b>
<b>konkrete Regelungen (ziviler Verkehr) IV1971</b>	<b>F:</b> les accords et décisions des quatre puissances au temps de la guerre et de l'après-guerre	<b>E:</b> legal position
<b>E:</b> detailed arrangements (civilian traffic)	<b>R:</b> соглашения и решения четырех держав военного и послевоенного времени	<b>F:</b> position juridique
<b>F:</b> arrangements détaillés (circulation de caractère civil)	<b>Ordnung, allgemein übliche Vorschriften bezüglich der öffentlichen IV1971</b>	<b>R:</b> правовая позиция
<b>R:</b> конкретные урегулирования	<b>E:</b> generally applicable regulations concerning public order	<b>Regelung, konkrete --en (ziviler Verkehr) IV1971</b>
<b>E:</b> bonded warehouse	<b>F:</b> les règlements généralement applicables en matière d'ordre public	<b>E:</b> detailed arrangements (civilian traffic)
<b>R:</b> консигнационный склад	<b>Organ, sonstige staatliche --e der BRD IV1971</b>	<b>F:</b> arrangements détaillés (circulation de caractère civil)
<b>E:</b> constituent part	<b>E:</b> other state bodies of the FRG	<b>R:</b> конкретные урегулирования
<b>F:</b> élément constitutif	<b>F:</b> d'autres organes étatiques de la RFA	<b>Regelung, nachfolgende Vereinbarungen und --en zwischen den zuständigen deutschen Behörden IV1971</b>
<b>R:</b> составная часть	<b>у учреждения государственной Республики Германии</b>	<b>E:</b> the consequent agreements and arrangements between the competent German authorities
<b>konsularisch, die --e Tätigkeit IV1971</b>	<b>Personen f pl IV1971</b>	<b>F:</b> les accords et arrangements entre les autorités allemandes compétentes qui s'ensuivent
<b>E:</b> consular activities	<b>Transitreisende identifiziert sind</b>	<b>R:</b> последующие соглашения и урегулирования компетентных немецких властей
<b>F:</b> les activités consulaires		<b>Reise, Gutscheine für --n und Rundreisen IV1971</b>
<b>R:</b> консульская деятельность		<b>E:</b> vouchers for travel and tours
		<b>F:</b> bons pour des voyages
		<b>R:</b> ваучеры для поездок и туров
		<b>Rundreise f IV1971</b>
		<b>E:</b> tour n
		<b>F:</b> voyage m

# THE TERMINOLOGY AND PHRASEOLOGY OF THE FOUR POWER ACCORD ON BERLIN

<b>Kontrollverfahren</b> n [V1971] E: inspection procedure F: procédure de contrôle R: процедура проверки	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры	E: permanent residents of -- F: les résidents permanents de -- R: постоянные жители	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры
<b>Kosten, Kompensation für Abgaben, Gebühren und andere</b> [V1971] E: compensation for fees and tolls and for other costs F: compensation pour les péages et taxes et pour les autres frais R: компенсация за сборы, пошлины и другие расходы	E: permanent residents of -- F: les résidents permanents de -- R: постоянные жители	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры
<b>Kriegszeit</b> f [V1971] Vereinbarungen und Beschlüsse der vier Mächte aus der Kriegs- und Nachkriegszeit E: wartime and post-war arrangements and decisions of the Four Powers F: les accords et décisions des quatre puissances au temps de la guerre et de l'après-guerre R: соглашения и решения четырех держав военного и послевоенного времени	E: permanent residents of -- F: les résidents permanents de -- R: постоянные жители	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры
<b>Luftfrachtdienste</b> p [V1971] E: air freight services F: service de fret R: услуги по воздушным перевозкам	E: permanent residents of -- F: les résidents permanents de -- R: постоянные жители	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры
<b>Macht</b> f [V1971] Vereinbarungen und Beschlüsse der vier Mächte aus der Kriegs- und Nachkriegszeit E: wartime and post-war arrangements and decisions of the Four Powers F: les accords et décisions des quatre puissances au temps de la guerre et de l'après-guerre R: соглашения и решения четырех держав военного и послевоенного времени	E: permanent residents of -- F: les résidents permanents de -- R: постоянные жители	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры
<b>Material, Verbreitung von -ien</b> [V1971] E: dissemination of material F: des objets à être diffusés R: материалы, предназначенные для распространения	E: permanent residents of -- F: les résidents permanents de -- R: постоянные жители	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры
<b>Mißbrauch der Transitwege</b> [V1971] E: misuse of the transit routes F: abus des voies de transit R: злоупотребление транзитных путей	E: permanent residents of -- F: les résidents permanents de -- R: постоянные жители	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры
<b>nachfolgende Vereinbarungen und Regelungen zwischen den zuständigen deutschen Behörden</b> [V1971] E: the consequent agreements and arrangements between the competent German authorities	E: permanent residents of -- F: les résidents permanents de -- R: постоянные жители	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры
<b>Rundreise, Gutscheine für Reisen und -n</b> [V1971] E: vouchers for travel and tours F: bons pour des voyages R: ваучеры для поездок и туров	E: permanent residents of -- F: les résidents permanents de -- R: постоянные жители	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры
<b>Schiene, Transitverkehr auf Straßen, -n und Wasserwegen</b> [V1971] E: transit traffic by road, rail and waterways F: circulation en transit par la route, le rail et la voie deau R: транзитное сообщение по шоссе, железным дорогам и водным путям	E: permanent residents of -- F: les résidents permanents de -- R: постоянные жители	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры
<b>Sicherheit, Angelegenheiten der - und des Status</b> [V1971] E: matters of security and status F: questions de sécurité et de statut R: вопросы безопасности и статуса	E: permanent residents of -- F: les résidents permanents de -- R: постоянные жители	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры
<b>sozialistisch, Büro der -en</b> E: matters of security and status F: questions de sécurité et de statut R: вопросы безопасности и статуса	E: permanent residents of -- F: les résidents permanents de -- R: постоянные жители	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры
<b>staatlich, sonstige -e Organe der BRD</b> [V1971] E: other state bodies of the FRG F: d'autres organes étatiques de la RFA R: другие государственные учреждения Федеративной Республики Германии	E: permanent residents of -- F: les résidents permanents de -- R: постоянные жители	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры
<b>Staatsgewalt, Akt in Ausübung unmittelbarer</b> [V1971] E: act in exercise of direct state authority F: acte accompli dans l'exercice d'une compétence étatique directe R: акт в осуществление непосредственной государственной власти	E: permanent residents of -- F: les résidents permanents de -- R: постоянные жители	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры
<b>ständige Verbindungsbehörde</b> [V1971] E: permanent liaison agency F: office de liaison permanent R: постоянный орган по связи	E: permanent residents of -- F: les résidents permanents de -- R: постоянные жители	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры
<b>Status, Angelegenheiten der Sicherheit und des</b> [V1971] E: matters of security and status F: questions de sécurité et de statut R: вопросы безопасности и статуса	E: permanent residents of -- F: les résidents permanents de -- R: постоянные жители	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры	E: persons identified as through travellers F: les personnes identifiées comme voyageurs directs R: лица, идентифицированные как транзитные пассажиры

# THE TERMINOLOGY AND PHRASEOLOGY OF THE FOUR POWER ACCORD ON BERLIN

<i>E: matters of security and status</i> <i>F: questions de sécurité et de statut</i> <i>R: вопросы безопасности и статуса</i> <b>Straße, Transitverkehr auf -n, Schienen und Wasserwegen [V1971]</b> <i>E: transit traffic by road, rail and waterways</i> <i>F: circulation en transit par la route, le rail et la voie d'eau</i> <i>R: транзитное сообщение по шоссе, железным дорогам и водным путям</i> <b>suspendieren v [V1971] die Bestimmungen des Grundgesetzes - sind suspendiert worden</b> <i>E: the provisions of the Basic Law - have been suspended</i> <i>F: les dispositions de la loi fondamentale - ont été suspendues</i> <i>R: положения основного закона - были приостановлены в своем действии</i>	<i>F: véhicules m pl</i> <i>R: перевозочные средства</i> <b>Transportmittel, individuelle (Personen) [V1971]</b> <i>E: individual vehicles (persons)</i> <i>F: des véhicules individuels (personnes)</i> <i>R: индивидуальные средства транспорта</i> <b>Transportmittel, vor der Abfahrt verplombt [V1971]</b> <i>E: conveyances sealed before departure</i> <i>F: des véhicules plombés avant le départ</i> <i>R: опломбированные перед отправлением перевозочные средства</i> <b>Übergangsstelle f [V1971]</b> <i>E: crossing point</i> <i>F: point de passage</i> <i>R: контрольно-пропускной пункт</i> <b>Verantwortlichkeit, Vermächte-Rechte und -en [V1971]</b> <i>E: quadripartite rights and responsibilities</i> <i>F: droits et responsabilités quadripartites</i> <i>R: четырехсторонние права и ответственности</i> <b>Verbindung, die -en der Westsektoren Berlins nach außen [V1971]</b> <i>E: the external communications of the Western Sectors of Berlin</i> <i>F: les communications des secteurs occidentaux de Berlin avec l'extérieur</i> <i>R: внешние связи западных секторов Берлина</i> <b>Verbindungsbehörde, ständige [V1971]</b> <i>E: permanent liaison agency</i> <i>F: office de liaison permanent</i> <i>R: постоянный орган по связи</i> <b>Verbindungsfunktionen f pl [V1971]</b> <i>E: liaison functions</i> <i>F: tâches de liaison</i> <i>R: функции по связям</i> <b>Verbindungsweg m [V1971]</b> <i>E: communication route</i> <i>F: voie de communication</i> <i>R: коммуникация f</i> <b>Verbreitung von Materialien [V1971]</b> <i>E: dissemination of material</i> <i>F: des objets à être diffusés</i> <i>R: материалы предназначенные для распространения</i>	<i>qui s'ensuivent</i> <i>R: последующие соглашения и урегулирования компетентных немецких властей</i> <b>Verfahren, festgelegtes [V1971]</b> <i>E: established procedure</i> <i>F: procédure établie</i> <i>R: установленная процедура</i> <b>Verfahren, geltendes [V1971]</b> <i>E: established procedure</i> <i>F: procédure établie</i> <i>R: установленная процедура</i> <b>Verfahren, Kontroll- [V1971]</b> <i>E: inspection procedure</i> <i>F: procédure de contrôle</i> <i>R: процедура проверки</i> <b>Verfassungs- oder Amtsakt [V1971]</b> <i>E: constitutional or official act</i> <i>F: acte constitutionnel ou officiel</i> <i>R: конституционный или должностной акт</i> <b>Verhältnis, das - zwischen den Westsektoren Berlins und der BRD [V1971]</b> <i>E: the relationship between the Western Sectors of Berlin and the FRG</i> <i>F: la relation entre les secteurs occidentaux de Berlin et la RFA</i> <i>R: отношение между западными секторами Берлина и Федеративной Республикой Германии</i> <b>Verkehr, Transit- [V1971]</b> <i>E: transit traffic</i> <i>F: circulation en transit</i> <i>R: транзитное сообщение</i> <b>verplombt, vor der Abfahrt -e Transportmittel [V1971]</b> <i>E: conveyances sealed before departure</i> <i>F: des véhicules plombés avant le départ</i> <i>R: опломбированные перед отправлением перевозочные средства</i> <b>Vermächte-Abkommen n [V1971]</b> <i>E: quadripartite agreement</i> <i>F: accord quadripartite</i> <i>R: четырехстороннее соглашение</i> <b>Vermächte-Rechte und -Verantwortlichkeiten [V1971]</b> <i>E: quadripartite rights and responsibilities</i> <i>F: droits et responsabilités quadripartites</i>
--	---	--

# THE TERMINOLOGY AND PHRASEOLOGY OF THE FOUR POWER ACCORD ON BERLIN

<b>Transitverkehr von zivilen Personen und Gütern</b> [V1971] E: transit traffic of civilian persons and goods F: la circulation en transit des personnes et marchandises civiles R: транзитное сообщение гражданских лиц и грузов	<b>Verdachtsgrund m</b> [V1971] Fälle, in denen hinreichende Verdachtsgründe dafür vorliegen, daß ... E: cases in which there is sufficient reason to suspect that ... F: des cas ... où il y aura raison suffisante de soupçonner que ... R: когда имеется достаточное основание подозревать, что ...	<b>R: четырехсторонние права и ответственности</b> <b>-e Wege</b> [V1971] E: routes designated for through traffic F: les voies affectées à la circulation directe R: по отведенным для сквозного проезда путям
<b>Behinderungen sein</b> [V1971] E: transit traffic will be ... unimpeded F: la circulation en transit ... ne sera pas entravée R: транзитное сообщение ... будет проходить без помех	<b>Vereinbarungen f pl</b> [V1971] E: wartime and post-war arrangements and decisions of the Four Powers F: les accords et décisions des quatre puissances au temps de la guerre et de l'après-guerre R: соглашения и решения четырех держав военного и послевоенного времени	<b>Vorschrift, allgemein übliche -en</b> bezüglich der öffentlichen Ordnung [V1971] E: generally applicable regulations concerning public order F: les règlements généralement applicables en matière d'ordre public R: обычно применяемые правила, касающиеся общественного порядка
<b>Transitweg m</b> [V1971] E: transit route F: voie de transit R: транзитный путь	<b>Abgaben für die Benutzung der -e</b> [V1971] E: individual tolls and fees for the use of the transit routes F: péages et taxes individuels pour l'usage des voies de transit R: индивидуальные пошлины и сборы за использование транзитных путей	<b>Wasserweg, Transitverkehr auf Straßen, Schienen und -en</b> [V1971] E: transit traffic by road, rail and waterways F: circulation en transit par la route, le rail et la voie d'eau R: транзитное сообщение по шоссе, железным дорогам и водным путям
<b>Transitweg, Mißbrauch der -e</b> [V1971] E: misuse of the transit routes F: abus des voies de transit R: злоупотребление транзитных путей	<b>Vereinbarungen, Ausdehnung von - und Abmachungen</b> [V1971] E: extension of agreements and arrangements F: extension d'accords et arrangements R: распространение соглашений и договоренностей	<b>Weg, für den Durchgangsverkehr vorgesehene -e</b> [V1971] E: routes designated for through traffic F: les voies affectées à la circulation directe R: по отведенным для сквозного проезда путям
<b>Transportmittel pl</b> [V1971] E: conveyances pl	<b>Regelungen zwischen den zuständigen deutschen Behörden</b> [V1971] E: the consequent agreements and arrangements between the competent German authorities F: les accords et arrangements entre les autorités allemandes compétentes	<b>Weg, Instandhaltung entsprechender -e</b> [V1971]

# THE TERMINOLOGY AND PHRASEOLOGY OF THE FOUR POWER ACCORD ON BERLIN

<p><b>E:</b> maintenance of adequate routes  <b>F:</b> l'entretien de voies -- adaptées à cette circulation  <b>E:</b> поддержание соответствующих путей  <b>Weg, vorgesehene -- (V1971)</b>  <b>E:</b> designated routes  <b>F:</b> les voies affectées  <b>R:</b> отведенные пути  <b>Westsektor, das Verhältnis zwischen den --en Berlins und der BRD (V1971)</b>  <b>E:</b> the relationship between the Western Sectors of Berlin and the FRG  <b>F:</b> la relation entre les secteurs occidentaux de Berlin et la RFA  <b>R:</b> отношение между западными секторами Берлина и Федеративной Республикой Германии  <b>Westsektor, die Bindungen zwischen den --en Berlins und der BRD werden aufrechterhalten (V1971)</b>  <b>E:</b> the ties between the Western Sectors of Berlin and the FRG will be maintained  <b>F:</b> les liens entre les secteurs occidentaux de Berlin et la RFA seront maintenus  <b>R:</b> связи между западными секторами Берлина и Федеративной Республикой Германии будут поддерживаться  <b>Westsektor, die Kommunikationen</b></p>	<p><b>zwischen den --en Berlins und Gebieten, die -- angrenzen (V1971)</b>  <b>E:</b> the communications between the Western Sectors of Berlin and areas bordering on --  <b>F:</b> les communications entre les secteurs occidentaux de Berlin et les régions limitrophes de ces secteurs  <b>R:</b> сообщения между западными секторами Берлина и пограничными с этими секторами районами  <b>Westsektor, die Verbindungen der --en Berlins nach außen (V1971)</b>  <b>E:</b> the external communications of the Western Sectors of Berlin  <b>F:</b> les communications des secteurs occidentaux de Berlin avec l'extérieur  <b>R:</b> внешние связи западных секторов Берлина  <b>Westsektor, direkte Durchreise von und nach den --en Berlins (V1971)</b>  <b>E:</b> direct travel to and from the Western Sectors of Berlin  <b>F:</b> le voyage direct à destination et en provenance des secteurs occidentaux de Berlin  <b>R:</b> прямой проезд в западные сектора Берлина и из них  <b>Wohnsitz, Personen mit ständigem -- in -- (V1971)</b>  <b>E:</b> permanent residents of --  <b>F:</b> les résidents permanents de --</p>	<p><b>R:</b> постоянные жители  <b>zivil, Transitverkehr von --en Personen und Gütern (V1971)</b>  <b>E:</b> transit traffic of civilian persons and goods  <b>F:</b> la circulation en transit des personnes et marchandises civiles  <b>R:</b> транзитное сообщение гражданских лиц и грузов  <b>zuständig, nachfolgende Vereinbarungen und Regelungen zwischen den --en deutschen Behörden (V1971)</b>  <b>E:</b> the consequent agreements and arrangements between the competent German authorities  <b>F:</b> les accords et arrangements entre les autorités allemandes compétentes qui s'ensuivent  <b>R:</b> последующие соглашения и урегулирования компетентных немецких властей  <b>Zuweisung von Bediensteten an das Generalkonsulat (V1971)</b>  <b>E:</b> assignment of personnel to the Consulate General  <b>F:</b> affectation de personnel au Consulat général  <b>R:</b> направление сотрудников на работу в Генеральное консульство</p>
---	---	--

8225

CSO:8320/0615



REVERSIBLE TRANSLITERATION OF CYRILLIC LETTERS

Munich NACHRICHTEN FUER DOKUMENTATION in German Vol 24 No 6, 1973 pp 239-241

[Special reprint of article by Joachim Schulz of Siemens AG: "Reversible Transliteration kyrillischer Buchstaben"]

[Text] Summary

In the evaluation and documentation of Soviet literature by means of electronic data processing (EDV), the cyrillic letters of the Russian alphabet must be represented by Latin ones in such a way that an unambiguous conversion is possible in both directions. The transliteration system established by the International Standardization Organization (ISO) is studied in light of this requirement. With the exception of one insignificant case, for which special treatment is necessary, the system proves to be suitable.

In evaluating technical and scientific literature from the Soviet Union, the problem arises of reproducing the cyrillic letters with Latin ones. This problem takes on particular significance when using electronic data processing (EDV) in the documentation.

It is just the use of EDV which has clearly pointed out that the problem of representing cyrillic letters has two sides: it must not only be possible to unambiguously replace the cyrillic alphabet with the characters of the Latin alphabet, but words and texts in this transliterated form must also go back into the original, i.e. cyrillic, form, and in fact, automatically. The conversion must thus be unambiguously reversible, i.e. "unambiguously unambiguous". Above and beyond this, such a transliteration should be as simple as possible to carry out, and nothing should stand in the way, even internationally, of its general dissemination.

Test of the ISO Recommendation

In contrast to a transliteration which has the goal of as precise a reproduction of the phonetic values as possible, thus, a transliteration which by nature must appear in different forms in the different languages, in the field

of documentation, recording and evaluation of fixed written data, the issue can only be one of a transformation true to the alphabet, thus, one of a transliteration. Since cyrillic script, more precisely, the Russian cyrillic alphabet as it applies today in the Soviet Union, has 33 letters and the German alphabet only 26, 27 or 30 (including umlauts and "ß"), it is not to be expected at the outset that one can make do in the transliteration with Latin letters alone. This means that special and diacritical marks will also be needed to reproduce cyrillic letters. Among the various possibilities for a transliteration, to be especially in the following is the system established in the ISO recommendation, No. 9 (1) as regards reversability.

Even if one works from the requirement of nonambiguity as the decisive criterion, thoroughly different transliteration systems are possible for the conversion of cyrillic to Latin letters. From the viewpoint of data processing, it would be enough to assign each cyrillic letter exactly one character of the Latin alphabet, incorporating various special characters (for example, \$, &, %) or numerals to be selected more or less arbitrarily, which may then no longer occur in their basic meaning in the transliterated text. With such a 1:1 conversion, reversible nonambiguity would be guaranteed.

#### Readability First

Another question, however, is the expediency of a documentation system as regards the users: the criterion of "readability" should in no case be neglected. Even if each cyrillic letter is assigned a corresponding Latin letter based on typographical form or even phonetic value, the additional utilization of arbitrary special characters for the remaining cyrillic letters would considerably impair the readability of texts transliterated in this fashion. A way out here is offered by reproducing cyrillic letters through the combination of several Latin letters (or other characters). Thus, for example, it would be possible to choose a representation for the cyrillic letters given in the table by numbers 8, 23, 25, 26, 27 and 31, in which by adding an "h" (which does not appear individually in the system) to a preceding basic letter, nonambiguity is assured (see the last column).

An additional possibility in contrast to such more or less arbitrarily constructable schemes, for which a standard form and generally binding force are hardly to be achieved, is the already mentioned ISO transliteration (see the table). This system, which has also been adopted in the German Industrial Standards (2), is based on the Czech alphabet. It has the advantage that it is internationally recognized and is also already used in German library transliteration with only a few deviations. (However, it is not unambiguous in both directions, since it does not take the hard sign into account; No. 28 in the table.) The ISO transliteration system uses only a few diacritical characters in addition to the basic letters of the Latin alphabet. In three cases (Nos. 27, 32, and 33), letter combinations are used in addition to these, and with two exceptions (Nos. 28 and 30) corresponding to each pair of upper and lower case cyrillic letters is one such pair of Latin letters.

# Transliteration of Cyrillic Letters

<u>Serial No.</u>	<u>Cyrillic</u>	<u>ISO</u>	<u>Arbitrarily Set Up for EDP</u>
1	а А	a A	a
2	б Б	b B	b
3	в В	v V	v
4	г Г	g G	g
5	д Д	d D	d
6	е Е	e E	e
7	ё Ё	ë È	
8	ж Ж	ž Ž	zh
9	з З	z Z	z
10	и И	i I	i
11	й Й	j J	j
12	к К	k K	k
13	л Л	l L	l
14	м М	m M	m
15	н Н	n N	n
16	о О	o O	o
17	п П	p P	p
18	р Р	r R	r
19	с С	s S	s
20	т Т	t T	t
21	у У	u U	u
22	ф Ф	f F	f
23	х Х	h H	kh
24	ц Ц	c C	c
25	ч Ч	č Č	ch
26	ш Ш	š Š	sh
27	щ Щ	šč ŠČ	shh
28	ъ Ъ	"	"
29	ы Ы	y Y	y
30	ь Ь	,	,
31	э Э	è È	eh
32	ю Ю	ju Ju	ju
33	я Я	ja Ja	ja

## A Linguistics Problem

The question of to what extent this transliteration is unambiguously reversible is now to be investigated in detail. For the transliteration of letters 1 to 5, 10, 12 to 18 and 20 to 23, nonambiguity is a matter of course, since a Latin letter corresponds precisely to the particular cyrillic letter here. Among the vowels, the Latin "e", "u" and "a" are utilized repeatedly, and among the consonants, the Latin "j", "c", "s" and "z". Furthermore, two more special characters or accents appear (Nos. 28 and 30). Viewed formally, i.e. when applied to any "series of characters" put together from the given "elements" of the cyrillic alphabet, such a transliteration is naturally not unambiguously reversible.

(The character "Ю" (No. 32), specifically "ju", when reversed could yield a "Ю" again or also the cyrillic letter combination "Ю" (Nos. 11 and 21).

However, such an approach does not do justice to the transliteration problem. The issue here is specifically not one of a purely logical and mathematical problem, but a linguistic problem. Thus, to be asked is which series of characters are then possible as words in the written language (taking into account the laws inherent in this language), and actually appear.

### Single or Multivalued

The above mentioned letters and their transliteration are to be treated individually from this point of view, in which case, working specifically from the Latin equivalents or the individual cyrillic letters, the question to be asked is one of whether the back transliteration is single or multivalued.

1. The Latin "e" is used to reproduce the "e" as well as the "ë" and the "э" (Nos. 6, 7 and 31); the two different diacritical marks, period and diaeresis, however, guarantee reversible nonambiguity: when a "e" appears, it is to be asked whether it is provided with a diacritical mark or not, and if so, with which one of the two (the same applies for "E").

2. The Latin "j" represents the cyrillic "ѣ", and in conjunction with the "u" or "a", the "Ю" and "Я" (Nos. 11, 32 and 33). Since the short "i" (No. 11) in Russian words does not appear before vowels (the letters "Ю", "Я", "e" and "ë" stand for the corresponding sound combinations) and also in foreign words only before the vowels "o" and "e" (3, 4, 5), an unambiguous decision can be made in the back-transliteration as to which cyrillic letter is involved: if an "a" or "u" follows the "j", then the result is an "Я" or "Ю"; otherwise, the "j" stands for an "ѣ".

3. The cyrillic hard and soft signs (Nos. 28 and 30) are to be represented by the characters ' and ', or ^ and ^^ (apostrophe and quotation marks or single acute and double acute accents). If both accents are used, they must be followed by a black space, so that on the typewriter, as well as on the corresponding teletypewriters, these diacritical marks do not appear over the following letters. When using apostrophes and quotation marks, it must naturally be assured that they do not occur in their actual function in the text to be transliterated. (The apostrophe is normally not used in Russian.) Otherwise, additional criteria would have to be set up to make a distinction.

Since the transliteration system for the hard and soft signs makes no distinction between upper and lower case letters, an unambiguous back transliteration seems to be placed in question on this point. However, the two characters never appear in Russian words at the beginning, so that an upper case writing is only possible if the entire word in which they appear is set in capital letters. This case can actually occur with automated electronic composition, so that the following decision is to be made. If a capital letter precedes and one follows, or when at least two capital letters precede, the corresponding cyrillic capital letters are to be used for letters Nos. 28 and 30.

4. The Latin letters "c", "s" and "z" occur both with and without a diacritical mark (we are dealing here with the little hook "ˇ"). The case of "ž" and "z" (Nos. 8 and 9) is to be solved based on the presence of the existing diacritical mark, just as in the case of the vowels "e", "ë" and "è". In the case of "č" and "š", it must be decided whether they are in the combination "šč" (No. 27) or not. In the latter case (Nos. 19, 24, 25 and 26), an unambiguous transliteration is again assured through the absence or presence of the diacritical mark. Only the case of the "šč" is more difficult, since in the case of a back-transliteration, this combination can yield both a "ш" (No. 27) as well as the combination of the two individual letters, "ш" and "ч" (Nos. 26 and 25).

#### A Few Exceptions

Normally the phonetic combination represented by "šč" is represented in modern Russian only by the letter "ш" (No. 27) or the letter combinations "сч" (Nos. 19 + 25) and "зч" (Nos. 9 + 25) (6). In a few very rare cases, however, the letter combination "шч" (Nos. 26 + 25) also occurs in modern Russian, which yields "šč" when transliterated. It follows from this that the back-transliteration of the character group "šč" must at one time lead to an "ш" (No. 27) and at another time to the combination "шч" (Nos. 26 + 25), and thus cannot be reversibly unambiguous. To be cited as an example here is the adjective "вешущатый" (from "вешушка"), in the formation of which the two letters "ш" and "ч" yield no "ш". (This assimilation takes place only when both belong to the same morpheme, thus, either the root or a suffix.) However, in the case here, the "к" in "вешушка" which becomes "ч" in the adjective through consonant alternation, is a component of the "ka" suffix, while the "ш" belongs to the root (7). Since this type of formation is only slightly productive (8, 9) and no other cases of a juxtaposition of the two Cyrillic letters are known to us, these few exceptional cases can be considered separately (and queried separately in an automatic transliteration with electronic data processing) so that the rule for the back-transliteration ("šč" yields "ш") remains valid. However, it appears to be more advantageous, also as regards as yet unknown or future word formations, to use a special distinctive or control character in such cases (perhaps a diacritical mark not otherwise used in Russian) to separate the two letters concerned. Other than its separating function, this character would have no other, and for this reason could be passed over when printing out transliterated texts.

#### The Unambiguously Reversible ISO System

With the exception of the case treated last above (No. 27), which will generally not occur because of its considerable rarity in broad areas, the transliteration based on the ISO system proves to be unambiguously reversible; for this reason, it is also fully applicable in electronic data processing, when the supplemental character introduced above is used. A technical prerequisite for this is actually the corresponding input and output equipment, which in addition to upper and lower case letters, as well as composition and special characters, also has a

corresponding stock of diacritical marks<sup>1</sup>. Where output equipment with cyrillic type is used, a text recorded and stored in transliterated form can be automatically, i.e. through the program, back-transliterated and put out in cyrillic type without further ado.

Address of the Author: Joachim Schulz  
Siemens AG, Dept. KB N FG SD 1  
8000 Munich 70, Hofmannstrasse 51;  
Home: 8000 Munich 71, Solothurner Strasse 70.

#### BIBLIOGRAPHY

1. ISO Recommendation R 9: International System for the Transliteration of Cyrillic Characters. First Edition, October 1955, Second Edition, September, 1968.
2. DIN 1460. "Transliteration slawischer kyrillischer Buchstaben" ["Transliteration of Slavic Cyrillic Letters"], October, 1962.
3. Shapiro A.B., "Russkoye Pravopisaniye" ["Russian Orthography"], Moscow, 1961, p 28.
4. Kopetskiy L.V., "Leksii po Fonetike i Morfologii Russkogo Yazyka" ["Lectures on the Phonetics and Morphology of the Russian Language"], Prague, 1965, p 17.
5. Tauscher E., E.-G. Kirschbaum, "Grammatik der russischen Sprache" ["Grammar of the Russian Language"], Berlin, 1960, p 31, footnote 1.
6. Chernykh P.I., "Historische Grammatik der russischen Sprache" ["Historical Grammar of the Russian Language"], German edition edited by H.H. Bielfeldt, Halle, 1957, p 137.
7. Shapiro, ibid., pp 95 ff.
8. "Grammatika Russkogo Yazyka" ["Grammar of the Russian Language"], USSR Academy of Sciences, Institut Russkogo Yazyka [Russian Language Institute], Two volumes, Moscow, 1960, v 1, p 333.

---

<sup>1</sup> The TEAM (= terminology recording and evaluation method) program system developed in the language service of Siemens AG, which works with this transliteration, uses six-channel perforated tapes for the data processing, which are produced on a type T-106 teletypewriter; available for the output are high speed printers with a corresponding stock of characters or electronic photo-composing systems of the DIGISET type of the Dr.-Ing Hell Co., Kiel.

9. Bielfeldt H.H., "Ruecklaeufiges Woerterbuch der russischen Sprache der Gegenwart" ["Reverse Dictionary of Modern Russian"], Berlin, 1958. (Three adjectives of this type are given).

Also:

Rozental' D.E., "Spravochnik po Pravopisaniyui Literaturnoy Pravke" ["Handbook on Orthography and Literary Proofreading"], Moscow, 1967.

8225

CSO:8320/0615

UDC 651.926:3:801.3

CONSIDERATIONS IN SETTING UP AND OPERATING TERMINOLOGY DATA BANKS AS A PREREQUISITE FOR MACHINE ASSISTED TRANSLATION

Munich NACHRICHTEN FUER DOKUMENTATION in German Vol 25 No 3, 1974

[Special reprint of the article by Karl-Heinz Brinkmann with Siemens, AG, Munich]

[Text] Summary

While the increasing flood of information is being talked about everywhere, only a small circle of experts seems to be aware of the problem of the flood of translation work closely related to it. Since we are still waiting for fully automatic, computerized language translation in a form which can be used in practice, the translation problem must be solved through machine assisted translation by means of terminology data banks. Consideration should be given to how efforts in this direction can be coordinated to avoid erroneous and parallel developments, and to assure compatibility of the various systems. The TEAM program system was conceived at the outset for working together with systems having similar goals.

At the end of 1972 in a Northern European capital, some famous publishers of common language dictionaries discussed the possibilities of using data banks. In this case, they came up with no draft plan of a data bank for the storage and utilization of a common language vocabulary, but all participants were agreed that the only alternatives in their technical field were between the use of data processing systems and retiring from the dictionary business.

If this is valid for the common languages and true of the goals set for dictionary production, then it applies even more to the quite dynamically developing technical languages, with an altogether more extensive spectrum of goals.

There is no doubt about the fact today that the growing mass of information in all fields of human knowledge and human activity cannot be recorded, ordered, opened up and organized for rapid access on the part of the interested parties without the use of the most modern means, and that means primarily, without the use of electronic data processing. Less well known to the



general public, but just as incontrovertible, is the fact that necessary to make information available and understandable, is, in the strictest sense of the word, its translation since a substantial part of the information appears in languages with which those interested are either not at all or inadequately conversant.

A start was made in the 1950's on the development of systems for fully automatic machine translation to find a solution to this problem. The results to this time make it appear doubtful whether a solution will ever be achieved in this field, which will make it possible to shift the information turnover to managed at points of language juncture from human translators to translating machines to any perceptible degree. In any case, we cannot wait for such a solution. The flood of information corresponds to an equivalent flood of translations. In managing them, the issue is not only one of a quantitative kind, but just as much or moreso also one of a qualitative problem (1, 2).

#### Constraining Factors in Translation

If one clearly understands this point, a check should be made as to what extent machine support can today contribute to the solution of the problem. For this, first to be established is which factors make the work of those active at language junctures difficult, predominantly that of the technical translator. They are probably largely well known, though nonetheless a few of them will be specifically cited again:

1. There is as yet no or no adequately clear terminology for the newest fields of research and technology.
2. Technical dictionaries cannot keep pace with development; they are almost without exception obsolete as soon as they appear.
3. Even the standardization of technical terminology is not at the state of the art in a timewise or quantitative sense.
4. The vocabulary, insofar as it can researched at all, is distributed over a multiplicity of sources of the most diverse quality.
5. There is a lack of a qualified supply of personnel, so that the existing translators generally work under the pressure of time considerations, and increasingly have less time to do terminological work.
6. There are only a few terminologists, those who pursue this profession and are active as such.

This listing is certainly not complete. If one is additionally clear on the point that a translator conscious of his responsibility, in light of the "tools" available to him, spends up to 60 percent of his time in researching, then one can formulate a number of desires which would provide an "ideal tool" for the translator: all information which he needs in his work should be provided with a minimum of effort on his part in the shortest possible time at an economically feasible price.

There is indeed no doubt that these desires can only be fulfilled by electronic data processing. The idea of bringing in electronic data processing to support translation activity is no longer quite so new. Even prior to the ALPAC 1966 report, which at once brought the developments directed towards machine translation to a halt and forced new conceptual premises on linguistics, procedures were known which attacked the translation deficit in a more practical sense. To be noted here in particular is the system of the at that time Translator Service of the Bundeswehr (today, Federal Language Office), which was pioneering in its own way, and the performance of which cannot be too highly evaluated if one keeps in mind the computer and programming engineering difficulties under which it was realized (3, 4, 5).

#### Developmental Coordination is Necessary

The further development of computer technology has made the creation of more modern, more convenient systems possible, such as that of the TERMIUM system of the University of Montreal (6) and the TEAM<sup>1</sup> system of the language service of Siemens AG, which is used as an example in the following (7, 8, 9, 10). It appears that similar projects are planned or have already been realized in many places. Such efforts are taking place at the most diverse levels: at large enterprises and state organizations, just as in the smaller company associations and translator unions. For this reason, it is certainly time to make an attempt to coordinate these efforts to a certain extent, so that expensive duplication and false developments are avoided. This article should make a contribution to this end and to encourage all of those engaged in related projects to make contact with each other. It certainly unrealistic now to believe that it is possible to set up a central data bank or at least a uniform data bank system. The individual interests can be brought under one roof only with difficulty, all those who are entitled cannot, and those with egotistical group, company or departmental interests, not at all!

#### Basic Rules for Cooperation

Still, the attempt should be made to formulate a few basic rules. In this case, the intention is not standardization. There should be no reproach if it is established here (with regret and resignation) that standardization in the field of electronic data processing and its applications has up to now remained strikingly behind developmentally speaking. In the case of interest to us here, a standardization in detail is not only unnecessary, but even undesirable. In order to make the digitally stored information in systems of this type interchangeable, and thereby make the systems compatible, only a few basic rules are to be observed:

1. Any kind of information loss must be avoided in the data recording. Such loss occurs when one works with simplified orthography, thus only with upper case letters and without accents. Commercially popular electronic data

---

<sup>1</sup> Terminology recording and evaluation method. The development of the TEAM system is supported by the Federal Ministry for Research and Technology.

processing has come to us from the English speaking world. This explains a great deal, but does not excuse the fact that in general so little use has been made of the great possibilities of character representation. One should not forget that among the languages which use the Latin alphabet, English with its complete lack of diacritical marks is not the rule, but rather an exception!

2. Information loss must also be avoided in subsequent processing. Such loss occurs if, in place of variable lengths for the individual information units, fixed ones are provided, and, as can later prove to be the case, they are of insufficient length. This frequently occurs because of a conceptual approach still oriented to the punched card and based on an old attachment.

3. The information to be stored must be broken down into the smallest still meaningful and clearly definable units, and these must be made individual accessible.

4. For the "information categories" arising in this way, a system must be set up which makes it possible to increase their number as required.

It is certainly evident that a free, lossless information exchange is possible either directly or via simple conversion programs between the data banks which meet these basic requirements, within the framework of an allied enterprise. If standardization is introduced, then it is introduced at such junctures.

The working data banks which are tied into each other do not basically have to contain the same information in all points. As regards the number of information types which can be recorded by them, they may not be subject to any limitations, and on the other hand, they can differ in the number of actually recorded types of information, since they will certainly frequently operate with different goals in mind.

#### Access to All Data Banks

It is questionable whether the transfer of large bodies of information from one data bank to another is basically worth striving for. In a united data bank system, it is sufficient if the users of all data banks have access to all other data banks. It is certainly to be preferred if special terminology stocks remain there where they were originally collected and probably can be best maintained.

There is yet one more very important point. The basic information, about which a lexical entry in a multiple language terminology data bank is structured, cannot be a term in any particular language. Moreover, one must work from concepts as the relevant meaning units. These are given designations in the various languages. Where there are no designations, or none as yet, or even only those which reproduce the conceptual contents incompletely, this circumstance must be indicated in a suitable fashion.

Grouped around the concept as the nucleus of the lexical entry of the data bank are the additional information units which are considered to be necessary or

useful, depending on the specific function of the data bank. Included here is naturally a recognition code which unmistakably differentiates the entry concerned from all other entries, so that one can correct, supplement or erase it at any time. Then there is data on the originator of the entry and the input date, primarily concerning the subject area though to which the concept applies.

#### Problematical Classification

This is now a quite problematical area, since in regard to the interchangeability and transferability of the information, one must naturally ask which system is to be used for the classification. If in the extreme case, each data bank in the system is operating with its own classification system, either the subject area code must be forwarded unchanged, something which means that all users of the combined system must know all of the classification systems used, or one must work with conversion programs at the interface points. In this case, there is actually then the danger that a loss of refinement or errors are produced in the conversion, because one classification method works with a different hierarchy or goes over the material with a finer comb than the other. There is one more possibility, specifically that one does not overly narrow the information categories for recording the subject area codes, so that the classification codes of different systems can be produced in it together. For example, with the TEAM system it is possible up to 30 subject area codes for an individual concept.

The greater part of the category system is reserved for the designations in the individual languages. Since there is considerable information which applies only to the specific designation or the particular language, the category system in this part is more expediently broken down into language blocks. In this case, the individual blocks, as well as the category system, should be expandable as a whole.

If it exists, a definition must be able to be given in each language. Where definitions are lacking, citations or contextual examples can be provided from the functional information sources. However, digital recording and storage of definition and source texts entail a considerable expense. Modern microfilm techniques offer interesting possibilities here. For example, one can transfer the definitions or corresponding texts from the original to microfilm, where they can be provided with codes which make their retrieval possible. These codes are also transferred to the digital data bank and provide for direct access to the filmed document. In modern microfilm retrieval systems, this access requires only a few seconds.

#### Text Reading Equipment is Still Lacking

The drawback to this analog storage consists at the present time in the fact that there is no equipment which can read any type of writing. Because of this, it is also not yet possible to conduct a computer search of texts recorded on microfilm in accordance with set criteria or to have them analyzed. If

this is desired for terminological work or linguistic purposes, they must be, now as before, digitally recorded and stored. The TEAM system operates at the present time in this fashion, however, is open for subsequent programming for a microfilm retrieval system.

Additionally offered as information units which are tied into the individual designations are: data on the parts of speech and other grammatical information, data on quality and authority, sources, data on the level of speech and the area over which it is disseminated, the phonetic transcription of the specific designation, semantic references, etc. Synonyms also belong here. The concept "synonym" should actually be quite narrowly defined, since every nuance of meaning, strictly speaking, already refers to another concept, to which an individual lexical entry logically should be devoted.

As already noted, the category system should be as flexible as possible, and admit of the subsequent introduction of additional categories. In the case of two-place numerical coding, 100 categories are possible per entry. TEAM operates at the present time with a system of this sort and provides ten categories in the so-called "entry heading" for general information. It then provides ten more categories in nine language blocks for the recording and processing of a maximum of nine languages per entry. This basic system has not yet been fully used up at the present time. If at some time it should no longer be adequate, a transition will be made to two-place alphanumeric coding, which would then permit an adequate number of categories for each individual entry in any case.

#### A High Degree of Freedom to Move Around

The processing of the individual information units of a terminological data bank requires so many different and demanding routines, that from the programming point of view, it would not be economical to allow all routines for all information categories. This would have to be done though if one really wants to achieve the unimpaired operational freedom in assigning categories, which is extolled by many manufacturers, but upon closer inspection is never available. For this reason, one should take such assertions cum grano salis. That a system developed for a terminological data bank can nonetheless offer a high degree of operational freedom and consequently, can also process information of a quite different character, has been demonstrated by TEAM through the application to telephone and book indexes, and especially through the production of registers.

The requirement for correct orthography was raised at the outset. It is now beyond doubt that this requirement can be met in any case in the storing of information. The same also applies to the data recording, although one could frequently get a different impression there. By having recourse to a coding with special characters, any desired character can be fed into the data processing system. This has also even worked for punched cards, although it was practiced there unwillingly. The more modern recording methods, for example, with a magnetic tape typewriter or via plain text readers, already provide for a greater stock of characters, but they still have to use special codings when working with diacritical marks. The drawbacks to this procedure are not as insignificant as one would like to make out. They consist in the fact that in

recording the plain text appearing on paper at the same time, which one absolutely needs if one wants to undertake proofreading and make corrections without having to first set up a computer run, involved are all of the shortcomings of the input, something which makes proofreading quite difficult. Linguistically versed people, who are really needed for proofreading, are seldom so "computer minded" that they would be ready to read large volumes of texts which have "been made so alien"! For this reason, TEAM still makes use of a special teletype which in addition to a perforated tape, provides a clearly readable record in unobjectionable orthography. It can represent all alphabets based on the Latin alphabet with all the special characters in the orthographically correct form, non-Latin alphabets in the manner recommended by the International Standards Organization (ISO), as well as all characters of the international phonetic alphabet in the simplest coding.

If one is not afraid of certain drawbacks, all known methods of recording are naturally permissible. In order to make allowances for them, a juncture is to be defined, i.e. the form in which the data from the recording methods are to be delivered to the data processing system is to be established. Once such a juncture exists, one can permit several recording locations with different recording methods for one data bank.

In any case, there is still no way of getting around the fact that the information to be stored in the data bank must be put in machine readable form through human effort. This is still the most time consuming and expensive bottleneck. For this reason, the volume of data which is to be matched to it, must be kept as small as possible. For the part of the information which is not to be digitally recorded, as already noted, microfilming is an option. In the case of the information to be digitally recorded, there is no question of duplicate recording for check purposes, as is frequently practiced in the case of punched cards, in light of the enormous volume. Flexible and convenient correction capabilities must appear for each phase of the recording and processing in place of this monitor capability.

#### Multiple Recording Must Be Avoided

Every other multiple recording of the information is also to be avoided. Thus, there is information which remains unchanged over an entire series of entries, for example, the subject area and source data, and the like. They should be recorded only a single time. In the TEAM system, they make up a so-called lead entry, after which only abbreviated entries are recorded. The input program produces the requisite complete entries from the lead and abbreviated entries.

Recording expense is also saved by the fact that synonyms are recorded in the entry of the associated concept. Instead of devoting their own complete entries to them during the recording, only where necessary is a reference or its own complete entry produced for each synonym by means of a program.

Something similar applies for multiple word designations with more than one word of significance. To look it up in the data bank and in alphabetical

lists, every significant word must appear once as a location criterion, thus, for example, as a word used in the alphabetization. This is achieved in the case of TEAM through a control character appended during the input, which serves to generate references or complete entries with a transposed word sequence. An alternative to this method which allows room for a certain subjectivity, would be an automatic "rotation" of the words, such as is used in a well-known English-German dictionary of data processing. In this procedure, each word of a multiple word designation (with the exception of certain trivial words) comes up in the first position one time. This leads, in the meantime, to a considerable inflation of the data stock. A crass, but in no way isolated example: "world trade telegraph double current line set expander" appears in the dictionary mentioned above in eight places!

### Considerable Savings

The savings which can be achieved in the recording, and in part also as regards the need for storage space, through utilizing the possibilities indicated above, are considerable. In the TEAM system, only about 50 percent of the characters, which can later actually be retrieved from the data bank, have to be recorded.

Which correction and updating procedures are to be provided during the input and during the later maintenance and servicing of the data bank, plays no direct part in their compatibility with other data banks and in their utilization. In the case of data banks which in design and structure correspond to the recommendations given here, the related problems can be solved in an optimum manner. Thus, at all processing stages in TEAM, information can be completely or partially corrected, erased, or added in any entry. If the same information elements in a series of entries have to be corrected, added or erased in the same fashion, this can be carried out with TEAM in one run with a single punched card for thousands of entries under certain conditions.

The requirements which were placed on the structuring and design of terminology data banks have only one purpose if they are viewed in relationship to the desired capabilities in the use of such systems. The most spectacular type of utilization is naturally direct interrogation of the data bank through a data viewing terminal. This has been technically possible for a long time now, but did not become actually economical until the introduction of operational systems for "virtual storage", which increase the total available working memory capacity of the systems to such an extent that a large number of participants with the most diverse desires can have proactively constant access to the information without having to incur disproportionately high costs for this.

### Direct Traffic with the Data Bank

Since there are already an entire series of information systems which make a direct man-computer dialogue possible, additional programming expenses can be avoided if the information of the terminology data bank is brought into the data stock of such a system through a simple conversion routine. Thus, the data stock of TEAM can be integrated into the GOLEM information system (11). Preferable where greater demands are to be satisfied is actually the case

where one can work directly with the terminology data bank. In this way, the user is offered far greater convenience, for example, the automatic initiation of partial and follow-up questions, if the answer to the original question is not satisfying. The possibility of obtaining spoken information (in the case of terminological information, probably preferably spelled out) from a data bank via telephone is only to mentioned here in passing.

The interrogation of the data bank can also consist in the input of question catalogs, which are answered through the output of high speed sprinter lists. The "textually referenced" or "textually synchronous" lists of the Federal Language Office have probably become the most well-known of the lists of this type, since they represent a true translation aid. TEAM also has this interrogation option, and a query can be directed from any of the recorded languages to any of the other recorded languages. If a multiple word designation or a phrase is not found as a whole, partial queries for the individual words can be made automatically. Furthermore, if the query is not answered, it is possible to generate follow-up questions for the overall question or parts of it, for which all entries are supplied which begin with the question or the specified part of the question. The result is thus a kind of word field output.

#### Computer Preparation of the Queries

The preparation of the questions in this procedure is naturally still accomplished by people. The further development of it sets for itself the goal of automatically extracting questions from texts which exist in machine readable form. This is certainly easier for some languages than for others, and depends on whether the languages are heavily inflected, whether multiple word designations can be isolated from the surrounding text, etc. The problem is not as difficult though as in the case of the automatic determination of the key words of a text which are relevant to the identification in the documentation, since for purposes of automatic interrogation of the terminology data bank, it is not harmful if words are also queried which do not belong to the technical language, and if homonyms are not recognized in the output text, and several possible translations are subsequently offered.

Not infrequently, a considerable expense must not infrequently be wasted on the translation of stereotype texts. Meant here are those texts such as appear in equipment descriptions, operating instructions, handbooks, servicing regulations, etc., again and again in completely equivalent or only slightly varying forms. It is obvious that it can be worth it to also store and process such texts in the data bank. Such texts are frequently changed only in small details, and under certain circumstances only for numerical data or the like. If the texts in all languages have the same breakdown, which is generally the case, the corresponding passages of the translation texts can be sought out by computer following correction or supplementation of the source texts, so as to then be correspondingly corrected or supplemented. In this way the today still considerable expense for updating such texts would be considerably reduced.



A utilization of the data bank with a final result which, with a superficial consideration, is conventionally quite attractive, is the printing out of glossaries and dictionaries in the broadest sense of the word. While glossaries are put out quite frequently on the high speed printer and serve a momentary need, so that, speaking purely superficially, they entail just a "hiccup from the computer", the products of computer controlled photo-composition of conventional dictionaries are no longer to be kept apart, for the fact is that they are checked thoroughly to see that they are up to date. For the production of a dictionary from a data bank with the modern means of data processing and photo-composition requires only a few machine hours.

#### Processing Capabilities of a Data Bank

The processing capabilities of a lexicographical utilization of a data bank which are desirable are to be enumerated in brief once again in this respect. First, there must be the capability of not only making selections from the data bank in accordance with each of the information categories provided, individually or in combination, but also in accordance with the contents or partial contents of the individual categories. The data stocks extracted from the data bank in this fashion must be supplemented by adding or transforming individual parts of the information, and be capable of being altered in accordance with the set goals. Complete or reference entries for synonyms, inverted multiple word designations, and abbreviations must be formed from the selected original entries. And finally, a provision must be made through special measures to see that in each language, in accordance with its own rules, information can be alphabetically sorted. This applies to the handling of umlauts in German and in other languages, as well as for "ch", "ll" and "ñ", or the transliterated cyrillic alphabet sequence, a, b, v, g, etc. Furthermore, the final purpose must be taken into account in the sorting, i.e. whether alphabetical sorting is to be accomplished throughout, or whether a paragraph or "nest" formation is desired. TEAM has all of these and more capabilities.

The TEAM system has a number of programs for the high speed printer output of the data stock prepared in this way. It is beyond the scope of this article to portray them here individually. All of these programs run on high speed printers with a large stock of characters, just as on those (with a corresponding automatic conversion) which have only the simple upper case alphabet of 26 letters. All programs can be used selectively both "on line" and "off line". By means of the new COM procedures, i.e. computer output on microfilm, it is also possible to put out the data prepared in this manner for high speed printer output via a COM unit on microfilm, instead of via the printer. This is carried out at a rate of around 70,000 characters/second and reduces the volume of the computer output to a fraction of a percent of the original paper volume. If terminology or similar data has to be made accessible at regular intervals to a large group of coworkers or interested parties, or even distributed world wide, such output on microfilm is to be recommended. Reading units are relatively inexpensive, and the investment is rapidly amortized through postage and freight savings, when instead of large quantities of paper, only microfilms are sent via letter.

With TEAM, the output is preferably accomplished via the Digiset CRT photo-composing system, in which the characters are electronically displayed on a cathode ray tube, and transferred by an optical system to a film or light sensitive paper, and if the occasion arises, also to microfilm. This output suitable for book printing is accomplished at far over a thousand characters/second. The text for this is prepared by the programs in such a way that it is recorded on a magnetic tape along with the requisite instructions for composition and breakdown, which then controls the digiset system. The first dictionary which was selected from a data bank and then processed in this way automatically to the point of a finished printing text with the pages made up, was the bilingual (German/English, English/German) "Dictionary of Data Processing" of Siemens AG which appeared for the 1970 book fair. The 400 pages of its two volumes were composed on the Digiset system in just short of one and a half hours from magnetic tapes, the production of which took less than a half an hour. The preliminary preparation of the data from the data bank lasted for about five machine hours.

As for the input, a juncture can also be defined at this point in the output from which point on a subsequent programming for any computer controlled composing procedure is possible. TEAM has such a juncture point.

#### TEAM Data Bank for a "Reading Course"

The possibilities for the use of terminology data banks are in no way exhausted by the examples cited up to now. In the Siemens company, the TEAM data bank is also used to determine the minimum vocabulary for so-called "reading courses" which must be mastered if the foreign language literature of a special subject area is to be understood when read. The duration of once such course is six to eight weeks for Russian.

The actual data bank program system can be rounded out by a series of auxiliary and supplemental programs. This has already taken place in the TEAM system, and is being pursued further. Important in this case are primarily programs which also support and facilitate the terminology work as the fundamental prerequisite for the structuring of the data bank. Included here are programs with which word and text concordances can be produced, as well as programs for backward alphabetical sorting and statistical investigations. This means that of interest for the terminology data bank are also programs of a type such as are used in linguistics. It can probably be said that the data banks which have appeared and are appearing in practice are doubtlessly of interest for linguistic studies in the technical language area, and not just in this other one. For a reservoir is available to linguists here which as regards volumes and manipulability leaves little more to be desired. It is to be hoped that the expense and effort of building up data banks based on practice in the terminological sector and for the procedures of machine assisted translation based on them, as well as machine assisted language training, can also well serve theoretical linguistics.

The question as to what extent the terminology data banks of the type described here, the development of which proceeded from the technical languages, can also

be employed in the common language sector, still remains open, as does that of whether the structures as well as the processing possibilities must be fundamentally changed for such an application. Research is underway in this regard at the present time, about which a report is to be made at the appropriate time.

Address of the author: Karl-Heinz Brinkmann, Siemens AG, Language Service,  
8000 Munich 70, Hofmannstrasse 51.

#### BIBLIOGRAPHY

1. "Language and Machines", Report by the Automatic Language Processing Advisory Committee (ALPAC), National Academy of Sciences, National Research Council, Publication 1416, Washington D.C., 1966.
2. Lehmann W.P., Stachowitz R., "Feasibility Study on Fully Automatic High Quality Translation", Final Technical Report, University of Texas, Linguistics Research Center, Austin, Texas, 1971.
3. "Dicautom - Consultation automatique de dictionnaires pour traducteurs humains" ["DICAUTOM - Automatic Dictionary Consultation for Human Translators"], LA TRADUCTION AUTOMATIQUE, Vol 4, No 3, 1963.
4. Schuck H.-J., "Grundzuege der maschinellen Uebersetzungshilfe" ["Principles of Machine Assisted Translation"], DER SPRACHMITTLER, Nos 1, 2, 3, 1965.
5. Krollman F., Schuck H.-J., Winkler U., "Herstellung textbezogener Fachwortlisten mit einem Digitalrechner - ein Verfahren der automatischen Uebersetzungshilfe" ["Producing Textually Referenced Glossaries with a Digital Computer - A Procedure in Automated Translation Assistance"], BEITRÄGE ZUR SPRACHKUNDE UND INFORMATIONSVERRARBEITUNG [CONTRIBUTIONS TO LINGUISTICS AND INFORMATION PROCESSING], No 5, 1965.
6. Dubuc R., "A Description of the Termium System of the Bank of Terminology [sic]", META JOURNAL DES TRADUCTEURS - Translators' Journal, Vol 17, No 4, Montreal, 1972.
7. Brinkmann K.-H., Schulz J., Tanke E. "Das Woerterbuch aus der Maschine" ["The Dictionary from the Computer"], DATA REPORT 4, Issue 4, 1969.
8. Brinkmann K.-H., Tanke E., "TEAM - Ein Programmsystem fuer die Loesung Terminologischer und Lexikographischer Aufgaben" ["TEAM - A Program System for the Solution of Terminological and Lexicographical Problems"], Siemens Publications Series "Data Praxis", No D 14/4059, 1969/1972.
9. Schulz J., "Le Système TEAM, une aide à la traduction" ["The TEAM System, an Aid for Translation"], papers from the international colloquium on linguistics and translation, META Journal des Traducteurs - Translators' Journal, Vol 16, Nos 1-2, Montreal, 1971.

10. Tanke E., "Das aktuelle Woerterbuch aus der Datenbank - Ein Beitrag zur Loesung sprachlicher Probleme im internationalen Informationsaustausch" "The Up-To-Date Dictionary from the Data Bank - A Contribution to the Solution of Language Problems in the International Information Exchange"], DEUTSCHER DRUCKER, No 38, 1971.
11. Hülck K, Klugmann D., Peetz G., "GOLEM - ein allgemein anwendbares Verfahren fuer die Dokumentation und das Wiederauffinden von Informationen" [GOLEM - A Generally Applicable Procedure for Documentation and Information Retrieval"], Siemens Publications Series "Data Praxis", No. 2-2600-48, 1969.

8225

CSO:8320/0615

UDC 800.935(083.72):82.03:681.3

# AUTOMATIC INTERROGATION OF A TERMINOLOGY DATA BANK<sup>1</sup>

Munich NACHRICHTEN FUER DOKUMENTATION in German Vol 27 No 1, 1976 pp 62-66

[Special reprint of the article by Joachim Schulz of Siemens AG, published by Verlag Dokumentation, ISSN 0027-7436]

## [Text] Summary

Multilingual terminology data banks are an important aid in understanding and translating technical texts. The automatic interrogation of stored terminology should identify the multiple term technical expressions in machine readable texts without human intervention, even in inflected form, and supply their equivalents in another language.

Data banks are used today in the most diverse areas of documentation and information. They have also proved themselves to be an effective aid in the solution of translation problems in science and engineering, the economy and politics. Central terminology data banks stored in a data processing system have been in existence for a few years in European institutions, at federal ministries and in industry. The value of such terminology collections depends not only on the stored information itself, but also quite substantially on the possibilities of making this information available in a suitable fashion in accordance with the requirements of all users, primarily the translator. Thus, interrogation procedures have been developed which extend from the directed output of stored terminology data via high speed printers or photo-composing systems in the form of complete dictionaries or selected or textually referenced technical glossaries to the direct answering of individual queries, for example via a data viewing terminal, thus in a dialogue between man and computer (1).

In almost all of the procedures for interrogation practiced up to now, it is still the task of the user to formulate the questions for the computer himself, i.e. to determine the technical expressions to be translated in the text and

<sup>1</sup> The work comprising the basis for this report was supported by funds of the federal minister for Research and Technology within the framework of the data processing program (Registration No. DV 5000). However, the author is solely responsible for the contents.

to put them in a certain basic form in which they can then be compared with the contents of the dictionary, the terminology data bank, stored in the computer.

Even years ago, some thought had been given to whether and how the translator could be freed from this work (2). If the text to be translated was already present in machine readable form (perhaps on composing perforated tapes on magnetic tapes), the program stored in the computer was to automatically prepare the text for possible questions, thus to extract the technical expressions contained in the text and supply their translation from the dictionary. Naturally, as a rule, the program does not know which expressions are unknown to the individual translator, and it will thus have to pose more questions and search for more answers under certain circumstances for a text, than the translator really needs in his work. A goal of this interrogation can be to automatically produce a glossary for a technical text which contains the technical expressions found in the text, along with their translations, or also a list in which the terms appear in the sequence of their occurrence in the text.

#### Primarily Two Difficulties

There are primarily two difficulties to be overcome during the corresponding processing of technical texts and the automatic translation of the technical words contained in them:

-- The individual technical words in the text can be inflected, i.e. appear in a form in which they are not stored in the dictionary, and on the other hand:

-- They can be part of a word group (a compound technical expression), within the framework of which they alone have their actual meaning.

Described in the following is a procedure (3) which on the basis of a set terminological data bank<sup>2</sup> makes do without extensive grammatical analyses, and is primarily set up for German and English technical texts, but can certainly also be extended to other languages.

It is a matter of course that the possibilities for interrogation and the search strategies to be employed depend substantially on the computerized technical dictionary which is used, from which, on one hand, the desired foreign language equivalents are to be supplied, but on the other hand, should also be used at the same time in the analysis of the texts to be processed. Before describing the procedure, the most important characteristics of the terminological data base used will be discussed in brief.

#### Terminology Data

---

<sup>2</sup> We are dealing here with the TEAM terminology data bank set up in the language service of Siemens AG, Munich.

-- The information units of the data base are multilingual terminology entries. The nucleus of each entry is a technical concept to be defined, which is provided with equivalent designations in the different languages (4).

-- Added to these designations, as in a good, printed technical dictionary, are supplementary information units: besides a definition, possibly contextual examples, subject area data, synonyms, source references, etc.

-- Based on the terminological arrangement and the extent to which the data base is multilingual, if precise or standard equivalents are lacking in a language, auxiliary expressions or paraphrases are also stored.

-- Technical designations can consist not only of individual words, but also of word groups (nouns with adjective, genitive or prepositional attributes, compounds in English, etc.). Thus, stored in the terminology data base are not actually words, but rather technical language, frequently complex designations, the translation of which does not as rule result from a 1:1 of the individual words contained in them. In the last dictionary derived from the data base, approximately one-fourth of the German and three-fourths of the English expressions were multiple word designations. (The question which is the basis of the technical concept formation and concept designation is to be distinguished from the question of idioms and phraseology, which is actually also relevant to the technical language (5).

-- The designations are stored in the usual basic dictionary form (as a rule, nominative singular, infinitive, etc.). Multiple word designations are recorded in their natural word sequence (for example, the German adjective in front of the noun). Important parts (individual words) of a multiple word designation which are not at the beginning, are especially marked during the data recording so that the entire designation can also be found via the corresponding inverted entries.

-- The designations are stored in true orthography, i.e. with upper and lower case letters, accents and other diacritical marks.

-- The entries and the partial information contained in them are arbitrarily long. A maximum length is in fact adopted for the machine processing, as a limit, but not reserved in the memory.

-- The number of information units, the dictionary entries, is practically unlimited. The terminology data stock far exceeds the vocabulary of the common language. In the field of electrical engineering with its allied areas alone, on figures on an order of magnitude of a million concepts.

-- The terminology entries contain only limited grammatical data. They form no special coded word form or root dictionary with all the data for the automatic grammatical identification and lemmatization of text words (6).

It is apparent that in light of the data volume mentioned above, as well as the continually necessary updating of a terminology data bank, a special preparation of this machine dictionary hardly appears realistic as regards

grammatical analyses. A corresponding linguistic (in addition to the terminological) preparation and maintenance could hardly be realized for reasons of personnel and time expense (at least at the present time).

Above and beyond this, it should not be forgotten that the existing terminology data banks serve a multiplicity of uses and as a consequence of the interrelationship with the already existing, entire data bank system, must be retained; besides the planned automatic interrogation and processing capabilities, all of the already realized ones must be preserved. If a change or adaptation of the existing data base is altogether necessary, it must be possible to execute it automatically, i.e. without expensive human intervention. This question will be treated in more detail in the following.

### Problems of Interrogation

The interrogation procedure is largely oriented to the reference work of the translator:

-- Text words, which in the given context are not trivial words, are sought in the alphabetically arranged dictionary.

-- During the search, thus in the comparison of a text word with the entries in the dictionaries, possible inflected endings are to be eliminated.

-- The identified text words are then immediately checked to see whether they are a part of a lexically listed word group in the given context.

Prior to the actual search in the dictionary, it is meaningful (as already indicated) to exclude all trivial words. What is to be considered a trivial word during the processing is to be established in light of a limited list, which is compiled based on frequency studies of corresponding technical texts, and can be rapidly searched through in the working memory of the computer.

The technical dictionary itself is to be stored only in a peripheral memory, for example, on magnetic discs, because of its scope, but can be interrogated there in a goal-directed, i.e. in a direct access mode. Searching in the computer dictionary means, just as in any other, the comparison of the queried word and the stored entries. In contrast to man, the machine can carry out this comparison only strictly according to the characters. Even slight differences can thus lead to failure in the interrogation. The questions of different ways of writing and the stock of characters used is not to be treated in any more detail here. It should only be noted that in connection with the difficulties cited for the interrogation, the differences between upper and lower case writing, as well as hyphens, gaps between words, and certain diacritical marks, are eliminated by the formation of an orthographical standard query in the computer. Among other things, this also permits a successful search for English compounds, regardless of whether they are written separately or together.



## Inflected Word Forms

In order find text words in a dictionary, they must, beyond this, be placed in the basic form applicable to the dictionary. Now there is no simple logical procedure here which permits making a distinction without extensive grammatical analyses based on formal graphemic criteria, whether one is dealing with an inflected word form in the case of an individual text word, and how this is to be traced back to the corresponding base form as the occasion requires.

If the automatic interrogation is to be carried out as far as possible solely with the information given in the data base, it appears to be expedient to work with an artificial reference form which one obtains through a reduction of the text words by dropping all possible inflected endings. The issue here is thus not one of a precise lemmatization; letter sequences at the end of a word are struck out, if they represent a possible inflected ending in the particular language without regard to whether such an ending is actually involved or not in the individual case. Endings are also dropped which are usually retained in the dictionary base form. The reduced word can then be found directly in the dictionary if a true ending which is suffixed to the dictionary base form was involved (Zugriffszeit/en [access time/s]), or if the character sequence struck out as the possible ending is in fact part of the base form, but as such was stricken out of the dictionary, i.e. in the corresponding indexing word of the dictionary entry (Prüfzeich/en).

## Procedures for Noun Expressions

Based on this formal treatment, the list of possible endings must yet be expanded by those character sequences which result from possible combinations of endings. In the specific case, thus, struck out along with a true ending is a preceding sequence of characters which are identical to a possible ending and as such are also suppressed in the reference form (Prüfzeich/en-s). Possible plural umlauts are also excluded (as are all others also) during the transformation into the orthographically standard form mentioned here (Bänder: Band). Likewise, the "ß" is generally resolved into "ss". Reference forms which are not clear to a person can result from the reduction, from the striking out of parts of the word root (for example, "Zin/s" in a manner analogous to "Gestein/s"), something which does not play any part though in the internal computer processing. In the first place, this procedure is naturally intended for noun expressions which predominate in the technical language. Separable verb compounds, for example, or also vowel gradation forms in the case of strong verbs, are not to be recognized in this fashion.

If all dictionary entries are supplied for a desired word, which differ from in only in the possible endings, one obtains superfluous answers under certain circumstances (for example, Funk/en, Funk/er, Funk/e for Funk). Only later experience can show whether these ambiguities resulting from the reduction described here in technical texts frequently occur, whether they can be accepted, or whether they must be eliminated through additional analysis effort (perhaps working from the paradigmatic compatibility of the various endings in the text word and in the dictionary entry).

## Reduction to the Internal Computer Reference Form

By making up a list of specific endings, a rule mechanism is defined for the computer by means of which the text words are reduced to the internal computer reference form. This rule mechanism is easy to modify and adapt to special lexical circumstances. Exceptions to the rules or limitations can be incorporated into the list. They are to be evaluated by the specific rule concerned. Thus, one can determine that the possible ending "s" is not struck out in German if it is preceded by "ni" or an "n" (along with "en") is only suppressed if it follows an "l" (Modul/n). If one catalogs the adjective ending "em" in the list, but renounces the possible ending combinations eme, emen, and ems, then the words Modem and System can be introduced as exceptions involved here. Of course, as regards the technical texts to be processed, also to be cataloged in the list of German endings are the endings of foreign words, such as ien, us, ius, etc. (Material/ien).

Under certain circumstances, it can also be advantageous to incorporate additional data on substitution endings into the rule mechanism, i.e. character sequences which in certain cases should replace another one. Thus, in English the ending "ies" could be replaced by a "y", when it is not expedient to generally replace all final "y's" by "i" except for those following a, e and o, and to strike out the ending "es". Above and beyond this, there is also the option of cataloging the "e" in the list of English endings, although it alone is no possible inflected ending, however, before an "s", can be a part of the plural ending.

## Multiple Word Designations

As has already been established, it is not enough to look up individual words in the dictionary and give their translation. Multiple word designations, thus expressions which consist of more than one word, are to be treated as lexical units. But there are difficulties in finding them as a whole in the dictionary, and not really only because individual parts of the multiple word designations can appear in the text in inflected form (. . . eines bistabil/en Multivibrator/s. . .). Above and beyond this, it is to be decided right at the outset prior to the interrogation and the dictionary search which word group in the text (of the many possible, i.e. beginning with which word and encompassing what number of sequential words) is to be posed as a question. The at times very great number of formally possible word chains (and thereby possible queries) within a sentence, could, for example, be reduced through a segmenting of the sentence by means of so-called stop words (2). The following approach appears to be promising in this respect, which proceeds in a manner analogous to reference to a conventional dictionary. The search is carried out for only the individual words, which are part of a multiple word designation, and as such, are stored in the dictionary with a reference to the complete designation. Thus, the dictionary itself should be used in the formulation of the queries.

It appears to be meaningful to limit oneself at the outset to the handling of fixed syntagmemes, such as represent the noun expressions, in addition

to the terms consisting of one word, without taking variable word and contextual structures (for example, verb groups) into account. Thus, multiple word designations are sought, the natural word sequence of which (as it is stored in the dictionary) is retained in the text. The condition for a successful search is to be restricted even more for these expressions: they are to be basically located through their first word. Then, the longest following word chain in the text for this initial word is to be sought, which is also stored in the dictionary (with the same initial word). The individual parts of a multiple word designation are generally subjected to the same inflection reduction as all other individual words.

### Textual Analysis

If the automatic interrogation is carried out in the manner sketched here to this point, it is possible to work up a given text word for word running from left to right. To accelerate the process, a list of the words most frequently sought in the dictionary and not found can be drawn up, along with the already mentioned trivial word list, in the working memory during the analysis. When the technical dictionary has reached a certain scope, in the case of these words we will be dealing more or less with words of the common language, which are then not specifically to be sought anew in the dictionary. Since no step by step reduction of the text words (possibly with an experimental generation of basic form endings) is carried out, in principle a relatively time consuming search process in the dictionary is sufficient, i.e. a mechanical disc access. A repeated looking up of information can be necessary only in the case of multiple word designations.

If one or more answer entries are found in the dictionary for a text word, apart from the already mentioned possible ambiguities of ending reduction, the following cases are to be distinguished:

-- A single word is involved which is taken as a technical word in the given text.

-- A so-called key word is present, i.e. a technical word which is stored in the dictionary as a part of a larger multiple word designation (for example, "failure" for "power failure": Netzausfall).

-- The text word is an initial word for one or more multiple word designations in the dictionary. Whether a multiple word designation and which one is present in the text is to be checked by means of the subsequent text words, in which case the longest word chain which is found in the dictionary as a whole (one begins with the greatest stored length) is to be taken as the pertinent technical expression (for example, "object" for "object module library", Modulbibliothek or for "object code", Maschinencode).

-- There is a combination of the above mentioned cases. Individual words are only put out as answers in this case if they are not part of a larger word group in the text. If direct agreement is recognized for an individual word, key words (and initial words) do not need to be put out as substitute answers.

It follows from what has been said up to this point that the interrogation procedure presupposes a correspondingly organized dictionary. Since the terminology data base stored on magnetic discs and used as the technical dictionary contains multilingual entries, where in principle each of the languages can serve as the source language, and furthermore, synonyms are permitted in the entries, several search concepts must as a rule exist for each dictionary entry. The access to an entry is thus expediently accomplished in two stages and the entire dictionary in the computer is correspondingly broken down into a root or a main part and an alphabetically arranged index for the main part. The index in which the first stage of the interrogation, the actual search takes place, is naturally again subdivided into the individual language sections and refers only to the root entries which contain the complete dictionary information.

Recorded in the index section are all search words for the dictionary entries, and in fact, expediently so in the standardized and reduced reference form described here. The index section is thus a result of the same automatic processing, or at least a part of it, to which the text to be analyzed is also subjected, and can thus automatically be built on to the data base by means of programs.

Search words are in the simplest case naturally all single word designations, after they have been put in the appropriate reference form. In the case of multiple word designations (not, on the contrary, in the case of mere re-writes), besides the initial words and key words possibly marked as such (as a location aid for the complex expression), the entire designations themselves are also to be recorded in the index in order to be able to more rapidly recognize the corresponding word group in the text. In this case, the components, and theoretically also only these, are to be reduced in accordance with the ending rules, where they could be inflected in their occurrence in the text. In German, as well as in French, standard groups are, for example, specifically those with the noun first and the possibly associated adjective; in English compounds, the last word is the noun, etc. The limitation on the reduction of individual terms of a multiple word designation would only be meaningful though if the terms concerned are also to be recognized in the text to be processed without greater effort. It is certainly simpler to basically subject each term of a multiple word designation to the same reduction in the formation of the total reference expression. Furthermore, it will probably be expedient to entirely suppress certain functional words in these designations in the formation of the search expression (*Zykluszeit des Speichers*, clearing of a call, terminal à écran). Each individual reference to an index entry is, among other things, provided with an indication which characterizes the relation to the corresponding dictionary entry. The totality of the index entries is automatically sorted and made available on a magnetic disc.

#### Translation Aids

Even if a text to be translated is not to be fully automatically processed in the manner depicted here, for example, as regards a glossary, but the intervention capabilities should remain open to the translator, this procedure can be helpful. It is conceivable and feasible to build this procedure

into a dialog system in which the human user is offered possible questions for the automatically prepared text for parts of it via display screen, in which case the translator accepts only the questions for which he desires answers, and at the same time still has the ability to "leaf through" the dictionary on his own for the offered questions. Even if the text is not available in machine readable form, the procedure can be used to expand the previous interrogation system in which the person still feeds the questions into the computer. This work can then be facilitated and the qualified translator relieved of it to the extent that the previously necessary reduction of the questions to the basic dictionary form can be eliminated and the entire sentence fragment underlined by the translator can be punched in without further processing and fed into the computer as a question.

Address of the Author:

Joachim Schulz, Siemens AG,  
Language Service, Hofmannstrasse 51,  
8000 Munich 70.

#### BIBLIOGRAPHY

1. Schulz J., "Der Computer hilft dem Uebersetzer. Ein Verfahren zur Abfrage eines auf Magnetband gespeicherten mehrsprachigen Fachwoerterbuchs" ["The Computer Helps the Translator. A Procedure for Interrogating a Multilingual Technical Dictionary Stored on Magnetic Tape"], LEBENDE SPRACHEN, 1975, 4, pp 99-103

Schulz J. and H. Goericke, "Das Woerterbuch im Computer. Moeglichkeiten der Direktabfrage eines mehrsprachigen Fachwoerterbuch ueber Terminals" ["The Dictionary in the Computer. Possibilities of Direct Interrogation of a Multilingual Technical Dictionary Through Terminals"], (Publication is planned in BABEL in 1976.)

2. Winkler U., "AUTOQUEST - ein Verfahren der automatische Abfragegenerierung" ["AUTOQUEST - A Procedure for Automatic Query Generation"], NACHR. DOK. 1, 1969, 1, pp 10-16.

3. See (1) and Tanke E., "Das aktuelle Woerterbuch aus der Datenbank. Ein Beitrag zur Loesung sprachlicher Probleme im internationalen Informationsaustausch" ["The Up-To-Date Dictionary from the Data Bank. A Contribution to the Solution of Language Problems in the International Information Exchange"], DEUTSCHER DRUCKER, 1971, 38, p 2-18.

Brinkmann K.-H., "Ueberlegungen zum Aufbau und Betrieb von Terminologie-Datenbanken als Voraussetzung der maschinenunterstuetzten Uebersetzung" ["Considerations in Structuring and Operating Terminology Data Banks as a Prerequisite for Machine Assisted Translation"], NACHR. DOK. 25, 1974, 3, pp 99-104.

4. Wüster E., "Die Struktur der sprachlichen Begriffswelt und ihre Darstellung in Woerterbuecher" ["The Structure of the Conceptual Language World and Its Representation in Dictionaries"], STUDIUM GENERALE, 12, 1959, pp 615-627.
5. Warner A., "Internationale Angleichung fachsprachlicher Wendungen der Elektrotechnik" ["Internationally Matching Technical Language Phrases in Electrical Engineering"], Supplement to ETZ, No 4, 1966.
6. Zimmermann H., "Das Lexikon in der maschinellen Sprachanalyse" ["The Lexicon in Computer Language Analysis"], Athenaum-Skripten Linguistik, Frankfurt am Main, 1972.

8225

CSO:8320/0615

THE LEXICOGRAPHY INFORMATION SYSTEM (LEXIS) OF THE BUNDESWEHR LANGUAGE SERVICE

Unknown DER SPRACHMITTLER in German No 3, 1976 pp 93-102

[Article by a Dr. F.B. in the information publication of the Language Service of the Bundeswehr]

[Text] The terminological work of the Federal Language Office is directed towards the requirements of the terminology users. These are primarily the translators; but language teachers are also drawing to an increasing extent on terminology collections for the preparation and execution of language training. Furthermore, professional people from the most diverse disciplines are numbered among the circle of users.

Terminology work in the Federal Language Office means the collection, processing and to a certain extent also the developing, as well as the reproduction of a multiple language vocabulary. All technical words in the German, English, French or Russian languages are recorded, where it is known that they have been used in translation or teaching assignments, or for which this can be anticipated. When I say "technical words", I mean the specific conceptually equivalent technical word pair in two of the cited languages, where German is the target language in the majority of cases. With the recorded technical words, the issue is both one of single term basic concepts, as well as one of multiple term concepts, frequently quite specialized terminology.

Applicable as regards a standardization of terminology is the fact that the German designations to be established, insofar as that is possible in the case of the technical concepts, should be simple, unambiguous and clear. They can be matched to international language use insofar as German language use does not contradict this. In accordance with the mission of the Federal Language Office, the recording of the actual language use ("ascertaining the actual norm") has predominance over the setting up of a theoretical norm.

The reproduction of the recorded and processed vocabulary makes up an important part of the terminology work, and one which becomes increasingly more important with the increasing vocabulary stock. The following considerations are the governing ones for the reproduction:

-- The success expected from the terminology work becomes greater the more the vocabulary is actually used by potential users.

-- A heavier use of the collected terminology leads not the least of all to an improvement in the vocabulary stocks, insofar as the terminology users return their experience and criticism to the terminology processor. The vocabulary is kept up to date really through use, through the daily interaction of the user with his material.

-- Language use is a common stock for each language, i.e. everyone has a share in it. This principle also applies to technical languages where trademarks are not involved. Copyright considerations as regards the use of association's own vocabulary stocks are directed against only reprints of entire technical glossaries which are identical in their content by third parties.

-- The reproduction of the vocabulary at public administration offices is, as a rule, done without cost within the framework of official assistance. The providing of terminology lists to interested parties outside the federal administration is based on the applicable regulations, i.e. costs are calculated.

The vocabulary recorded in the Federal Language Office and in a few other places is stored and processed electronically almost exclusively within the framework of computer lexicography. A central "double bookkeeping" with conventional index cards is not provided because of the high personnel expense this entails, even if index cards enjoy increasing popularity as source materials among terminology workers and terminology users. In the Federal Language Office, index cards are used for only a few special stocks, in which facilitate the handling of the vocabulary. The breakdown of the word locations recorded in LEXIS can be clearly explained by means of one such index card:

The individual blanks in this card correspond to the structure of the data entries and are to be filled out as follows:

I <sub>1</sub>	GAINAGE			40
				35
I <sub>2</sub>	/MATIERE DE = POUR CABLES			34
				41
I <sub>4</sub>	KABELMANTELMISCHUNG			45
				30
I <sub>5</sub>	(DIN 7730, 1965)			39
				36
I <sub>3</sub>	B2 F11	RA49	+	



(Source language or target language)

- I<sub>5</sub> : Explanatory addition or transposition in German
- I<sub>3</sub> : Positions 1 and 2 : Language or language direction symbol  
Positions 3 to 5 : Subject area code, special code  
Positions 6 to 11 : Source code  
Position 12 : Quality symbol for the foreign language expression  
Position 13 : Quality symbol for the German expression.

Each line corresponds to an available position in these data fields. The number of positions is specified with numerals. Information in the second line is placed in a so-called excess length stock.

The large field at the bottom on the card should record additional data (context, definitions) which in the future could be recorded in a background store. The small fields at the right bottom are for notes of the lexicographer (references to transpositions, committee meetings), as well as for specifying the processor or examiner, and the date.

The vocabulary stored in the computer center can be printed out either in part or entirely by the computer. For reasons of expediency and economy, particular, individual vocabulary stocks put together in subject areas are taken out of the total vocabulary stock in accordance with the desires of the users. As regards the output format, there are the following publishing capabilities:

1. Microplan film (format 105 x 148 mm, reduction 48:1); output by means of COM systems\*. Table readers with a 150 x 200 mm screen (magnification, 1:40) are available for reading microfilm dictionaries in our language service. The microplan films can be rapidly and economically produced and are planned for issuance anew to all users semiannually.
2. Computer tabulation sheet (DIN A 3, continuous form).
3. Reproduction from computer printout reduced to DIN A 4 in an offset printing process, or as XEROX copies.
4. Printing in a photo-composing process.
5. Display output on a data viewing terminal. This capability is, within the scope of remote data processing, limited at the present time to four locations in the Cologne - Bonn area.

The entire multilingual vocabulary stored in the computer center and classified in accordance with the language directions and subject areas (around half a million bilingual, different word locations) was printed out in years past almost exclusively in the form of computer tabulations. These word lists

---

\* COM = Computer output on microfilm.

are expensive and time consuming in their preparation and shipping, as well in the large amount of relative inconvenience in their use. For this reason most translators could not obtain the technical glossaries in the desired scope.

By means of microfilming electronically stored vocabulary stocks, it is possible with a relatively small outlay to produce microfilm dictionaries on microplan films (microfiche) (48-fold photograph reduction). One microplan film in the DIN A 6 format contains around 15,000 data entries. Previously needed for this were more than 200 pages of a DIN A 3 computer tabulation.

Thus, the output of word indexes has the following advantages:

- Lower production costs than with previous procedures
- Less expense for procurement and shipping
- Easier handling by the users.

Data processing within the scope of computer lexicography serves several purposes:

1. It provides procedures for the production of monolingual and multilingual dictionaries thanks to the manifold and efficient output capabilities. The short term compilation of technical glossaries which can be combined in different ways in accordance with the wishes of a user is to be considered a special advantage. About 500 such assignments are processed annually in the Federal Language Office, which yield about 300,000 pages of technical glossaries. Not only translators make use of the lists. Professional people are also taking advantage of these lists to an increasing extent, where they go abroad for courses in a particular technical field.
2. It permits the preparation of textually reference technical glossaries within the framework of the translation procedure supported by data processing, in particular for the large projects of a translator team. The textually referenced technical glossaries, as is well-known, serve the dual purpose of facilitating the work of the translator and making a check of the correctness and completeness of the stock possible. In other words: the textually referenced technical glossary has proven itself to be, primarily in the past, a very good means of increasing the stock. Then the first textually referenced list was printed out in 1966, the data encompassed 110,000 entries. No one is surprised that at that time over 70 percent of the queried word entries (English) yielded a 'information lacking' indication, but just because of this, the Federal Language Office was able in four years to more than double the English-German stock with the help of the users and the textually referenced technical glossaries. Today, the percentage of missing information is far less and the textually referenced lists derived from the translation assignments now supply only 15,000 entries annually. But this figure is also significant if one considers that over 40,000 word locations are stored in the subject area of electrical engineering and electronics alone.

3. It supplies lexicographical aids such as code registers, key word catalog, and source index.

4. It is an indispensable aid for lexicographical work, especially because it takes over the following tasks:

- Rearranging a vocabulary stock (for example, output in the reverse language direction).
- Aggregate changes (code groups, character sequences in word locations).
- Doublet search in the stock (for searching out and erasing quasi-identical data entries).
- Checking for doublets during input (any entry which is identical with an already existing entry is rejected).
- Individual lexicographical investigations by means of concordances.
- Preparing statistics (sentence counts in accordance with the first character of the entries (in order to be able to break the vocabulary down into equal sized parts during publication and editing), as well as in accordance with subject areas/special stocks).
- Automatic monitoring of individual data fields during data recording (plausibility checks). This computer check process facilitates and reduces the considerable work effort in proofreading.
- Documentation of the entire emendation service.

A prerequisite for these manifold application possibilities was the creation of an extensive user program (around 20 TP<sup>1</sup> programs as well as about 50 programs for batch processing) for the foreign language work. Furthermore, about 20 IMS<sup>2</sup> auxiliary programs for operational run security, as well as reorganization and loading of the data banks, were required.

The total vocabulary (about 900,000 data entries) is stored and processed in Bonn. Located at the Federal Language Office in Hürth and a few other locations are the data transmission facilities and data terminals needed for an efficient operation (modem, displays, printers).

The LEXIS system of the language service of the Bundeswehr is primarily set up for the requirements of practice and is thereby close to the user. The user is incorporated into the circuit, he is a part of the system, and in fact a very high speed system. The terminology information is referenced to its salient contents. Some things were consciously omitted in Phase I of the system (basic procedure) which many specialists will perhaps miss. Thus, for example, there are no diacritical marks and there is only the standard upper case letter writing style. The compression of the information (as is

---

<sup>1</sup> TP = Teleprocessing

<sup>2</sup> IMS = Information management system.

anticipated by a trained technical translator) is faced with omitting hierarchy, relationships, especially with a homonym and synonym reference system, and also with grammatical information such as data on gender, etc. Each word location stands by itself. The context and definitions were previously also not as a rule recorded. Additionally capabilities will come into play here with the further expansion of the system in Phase II (see the appendix). On the other hand, the system is very flexible. The storage of special stocks (as for a special translation project) is possible without any difficulty. Thanks to the scope achieved in the interim, the level of coverage is considerable, even if different from language and from subject area to subject area; in many technical fields, "hit percentages" of 80 percent and more are achieved with textually referenced queries.

LEXIS is as much a system for producing dictionaries as an efficient interrogation and reference system for the trained linguist, can be updated quickly and has an efficient input and correction capability. In this way, it meets the needs and requirements of the user to a considerable extent.

#### APPENDIX

##### Expansion Phases of the LEXIS System

###### Phase I

Basic procedure (disc oriented, using TP and an IMS)

\* On-line data recording of:

- Entries;
- Rejections;
- Textually reference queries.

\* Emendation service (weekly) with:

- Doublet search;
- Check for formal errors;
- Check for authorization for intervention in the stock;
- Cumulative supplements for the individual users;

\* Processing of the textually referenced queries:

- On-line recording (see above);
- Processing in batch operation;
- Return transmission of the results via TP;
  - Output in: Textually referenced sequence
  - Alphabetical sequence;

\* On-line stock interrogation (dialog traffic):

- Query for identity;
- Query for "equal to and greater";
- Paging through;

-- Directions: Foreign language - German  
German - Foreign language  
Register number

\* Stock servicing on-line

-- Changing entries;  
-- Duplicating entries with or without simultaneous changing (for example, to transfer an existing entry into an additional subject area).

\* Batch processing

-- Query for language/language direction, subject areas and/or sources;

Types of output: High-speed printer  
Photo-composer  
COM microplan films (microfiche)

-- Concordances  
-- Aggregate erasure or modification of code groups;  
-- Aggregate erasure or modification of character sequences in the actual dictionary entries;  
-- Statistics.

Phase II

Expansion of the Basic Procedure

\* Setting up a background store for recording:

-- Excessively long word locations;  
-- Definition texts or contextual examples;  
-- Source citations;  
-- Cross-references to subordinated, superordinated and associated concepts, including:  
-- Synonyms/antonyms;  
-- A register of entries with transposed elements in their normal word sequence.

\* Expansion of the textually referenced and on-line stock interrogation

-- Principle: "Similarity instead of identity" between the queried character sequence and the store contents;

\* Modified doublet search:

-- Search for "equal" entries in the stock, neglecting individual parts of the sentence structure ("quasi-doublets").

\* Auxiliary programs to facilitate the supervision of the stock:

-- Setting up archives for the results of textually referenced, subject area referenced or concordance queries, and comparison with a later query of the same specification, in order to recognize changes in the memory which have entered in the interim.

-- The input of the results of subject area referenced queries as textually referenced queries.

#### Some 1.8 Million Textually Referenced Queries in 10 Years

The first run of about 2,000 textually referenced queries was processed on June 29th, 1966, in the Bundeswehr Trier computer center, and the result turned over to translators of the, at that time, Translation Service of the Bundeswehr in Mannheim in the form of tabular forms. Until today, thus in a space of about ten years, about 1.8 million foreign language and German expressions have been processed in this form. The "dictionary" stored at that time had about 110,000 entries; today the stock of bilingual dictionary entries (primarily also thanks to the indefatigable cooperation of translators and proofreaders) has grown to about 900,000.

The "percentage of hits" increased, depending on subject, from first around 30 to 40 percent to 60 to 70 percent. While ten years ago the queries were still fed in with perforated tapes, later, following conversion of the computer vocabulary processing on another electronic data processing system, with perforated cards via a remote data link, today one is served in the Federal Language Office by modern display terminals, which are directly connected to the central computer.

The textually referenced query procedure has basically not changed, despite various technical improvements in its particulars. Then, as is well known, it was for a long time the first practicable procedure of this type; today, more than a half a dozen procedures are in operation.

As one of several possibilities for information acquisition within the framework of the vocabulary processing system, it has passed its performance trial, especially in the handling of large scale translation projects of the translation service of the Bundeswehr and the Federal Language Office, on which a multiplicity of translators worked at the same time. Without this support, the terminological coordination of their work on such a project would have been almost impossible.

The wait for computerized translation procedures, the hope for which was still held out 10 years ago in many places, could not go on much longer, and the language service of the Bundeswehr, skeptical of this from the outset, limited itself in the processing of a bilingual vocabulary to electronic data processing applications which could be realized and be of service in practice. In this way, a development was initiated from which practically all translators of the Bundeswehr profit today.

Dr. F.B.

8225

CS0:8320/0615

END

---